

# LEAD SCORE CASE STUDY

---



# 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.

ALTHOUGH X EDUCATION GETS A LOT 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

X EDUCATION HAS APPOINTED YOU TO HELP THEM SELECT THE MOST PROMISING LEADS, I.E. THE LEADS THAT ARE MOST LIKELY TO CONVERT INTO PAYING CUSTOMERS.



## BUSINESS GOAL :

Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.

A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted

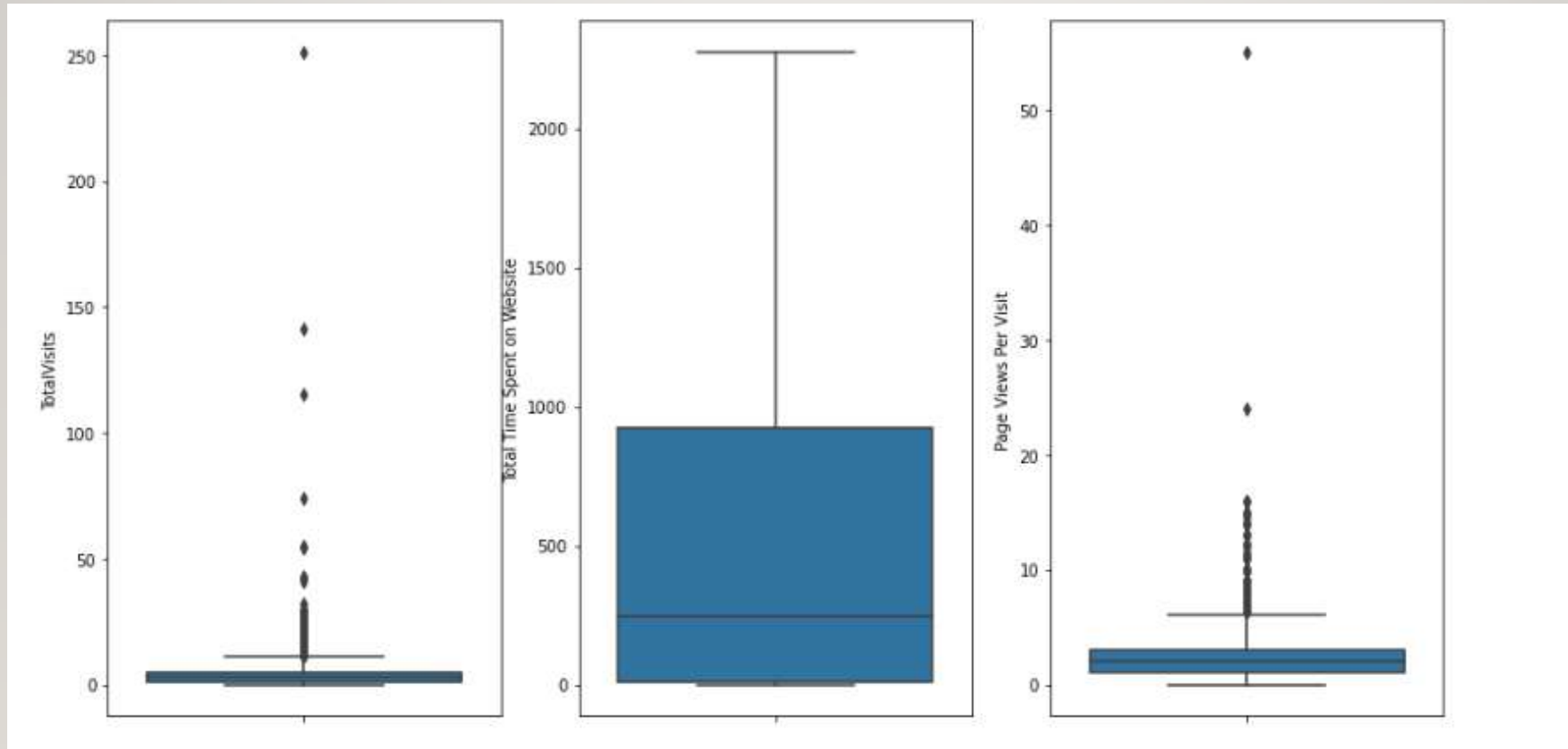


# STRATEGY

- Data import
- Data Cleaning
- EDA Analysis
- Test Train Split
- Scaling
- Model Building
- Model Evaluation by different metrics and measures
- Model Predictions on Test Data Set

## OUTLIERS:

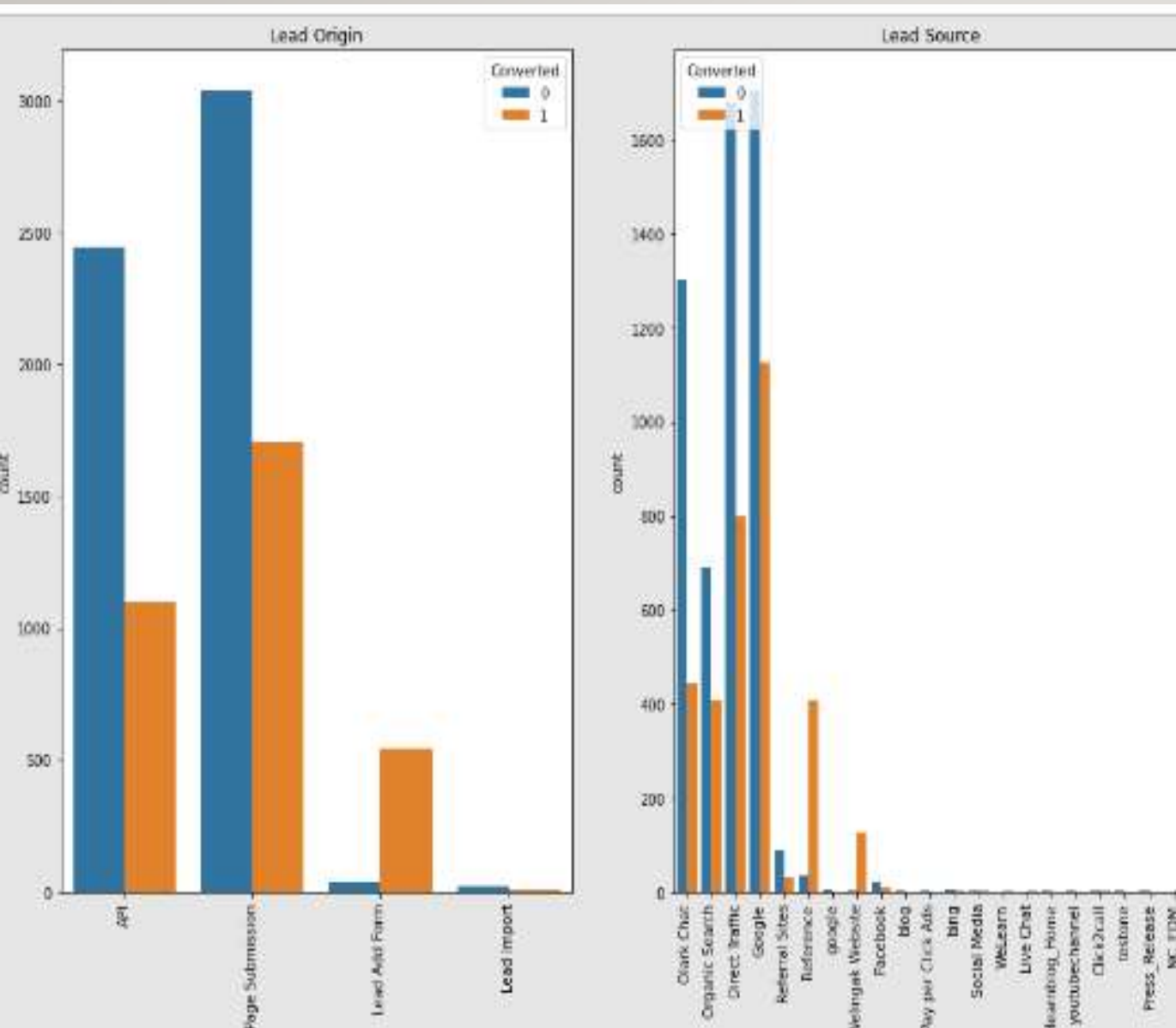
Finding outliers in total visits, total time spent on website & page views per visit  
And we find outliers in page views per visit



# Exploratory Data Analysis





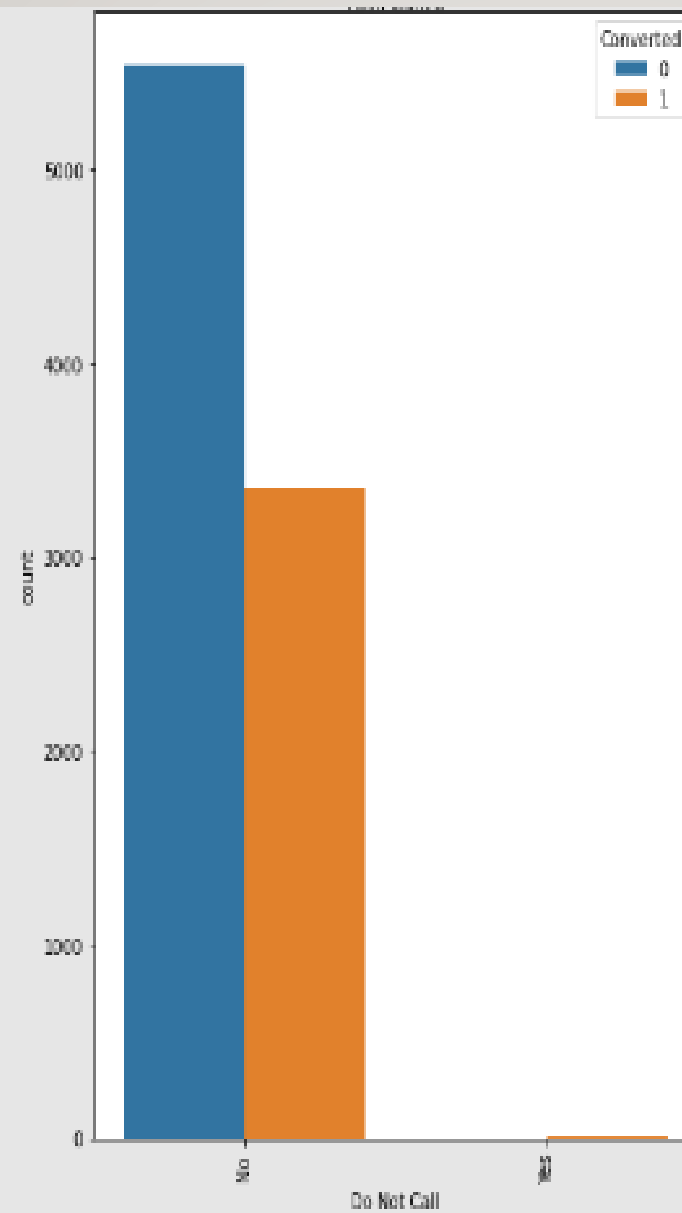
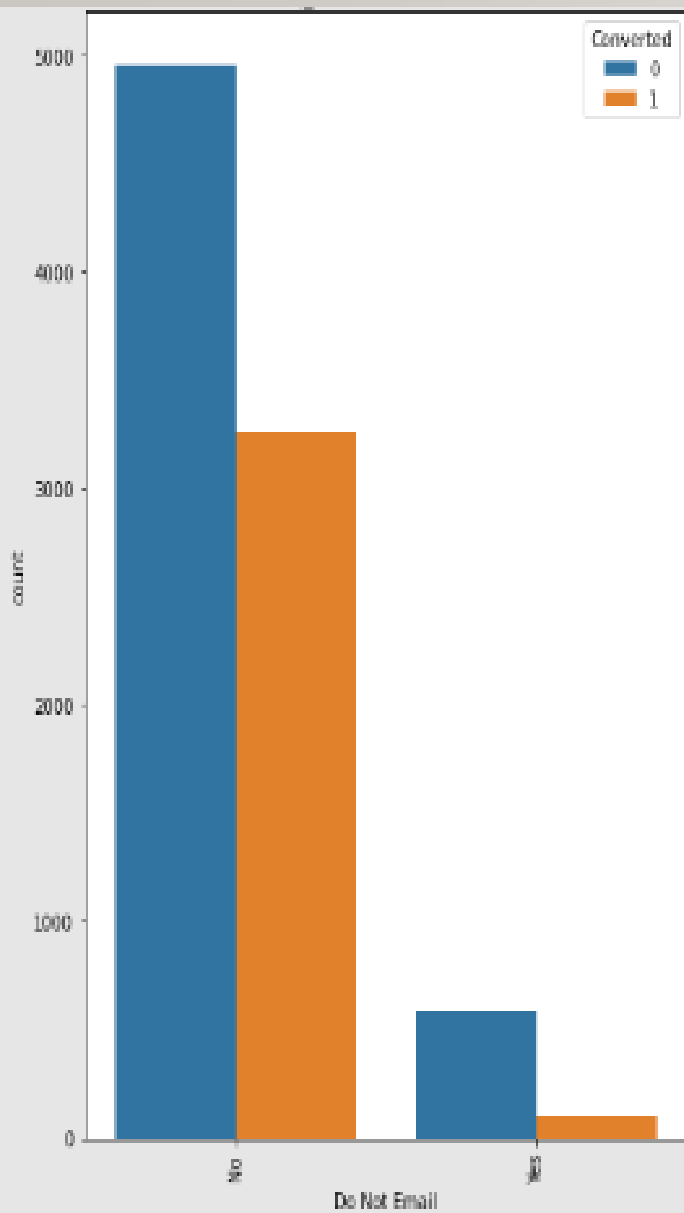


**Lead Origin Vs Converted :**

Landing page submissions has had high lead conversions

**Lead Source vs Converted :**

Google Searches has had high conversions compared to other

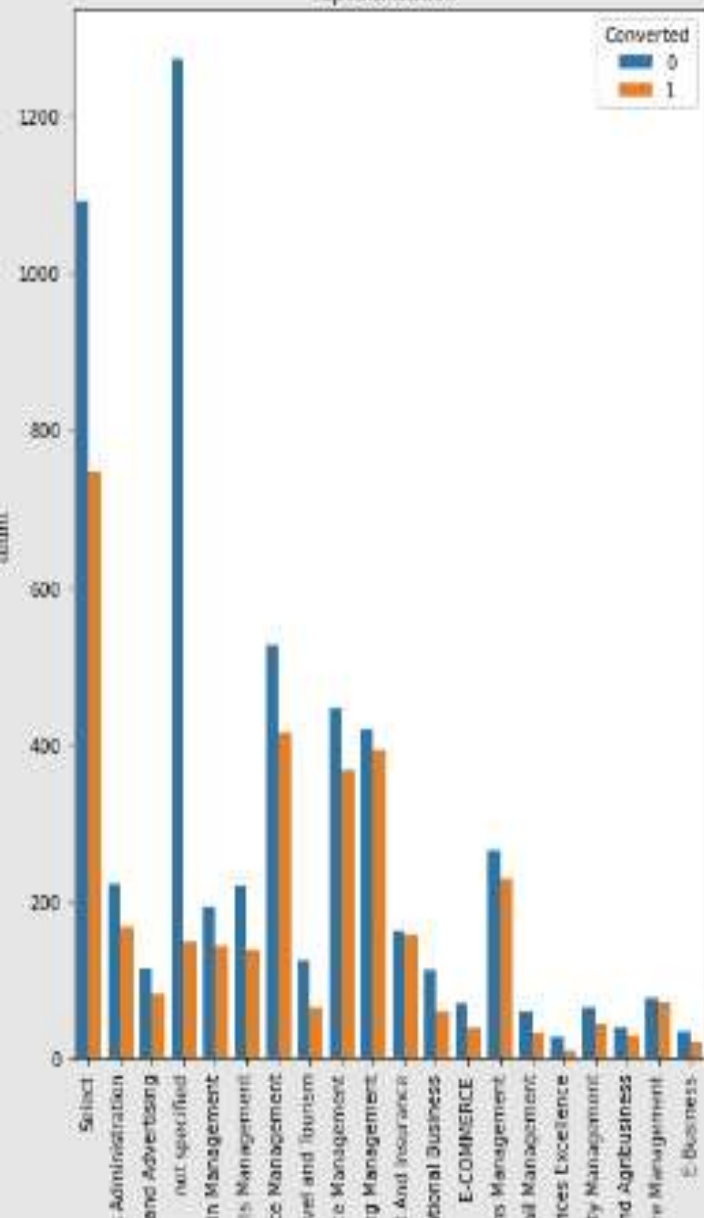


**Do not email vs Converted:**  
Google searches has high conversations

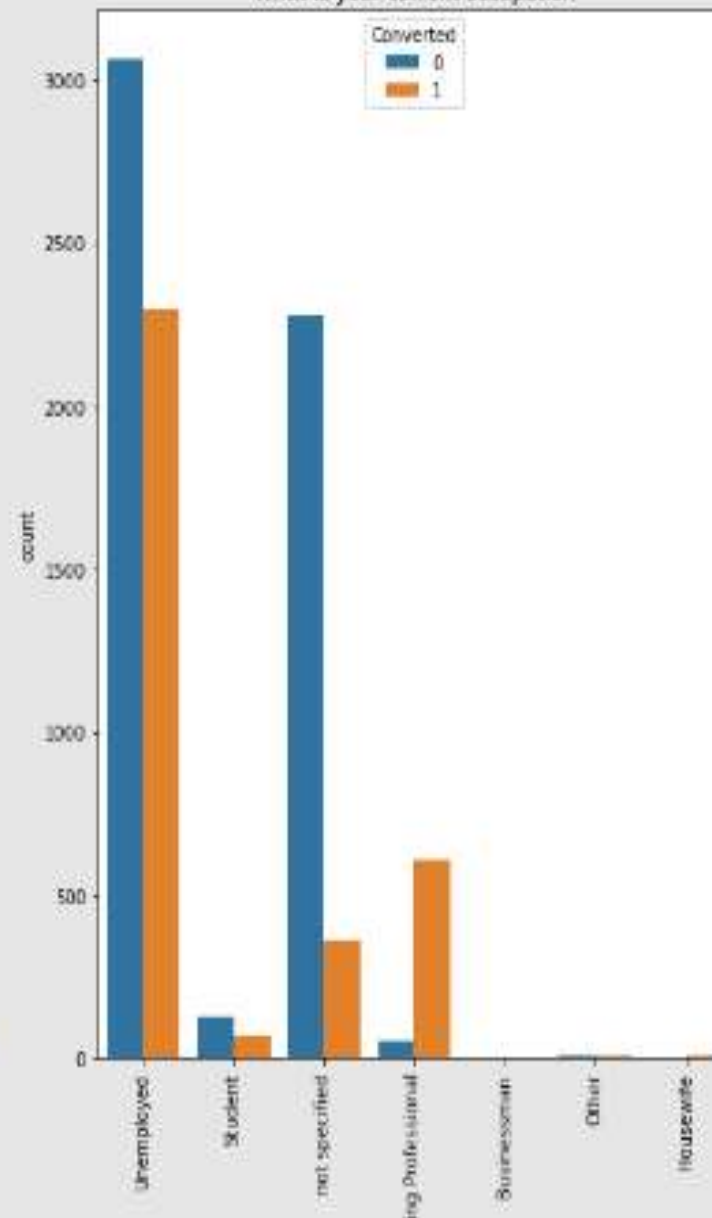
**Do not Call vs Converted:**  
Most leads are not informed via phone



Specialization



What is your current occupation

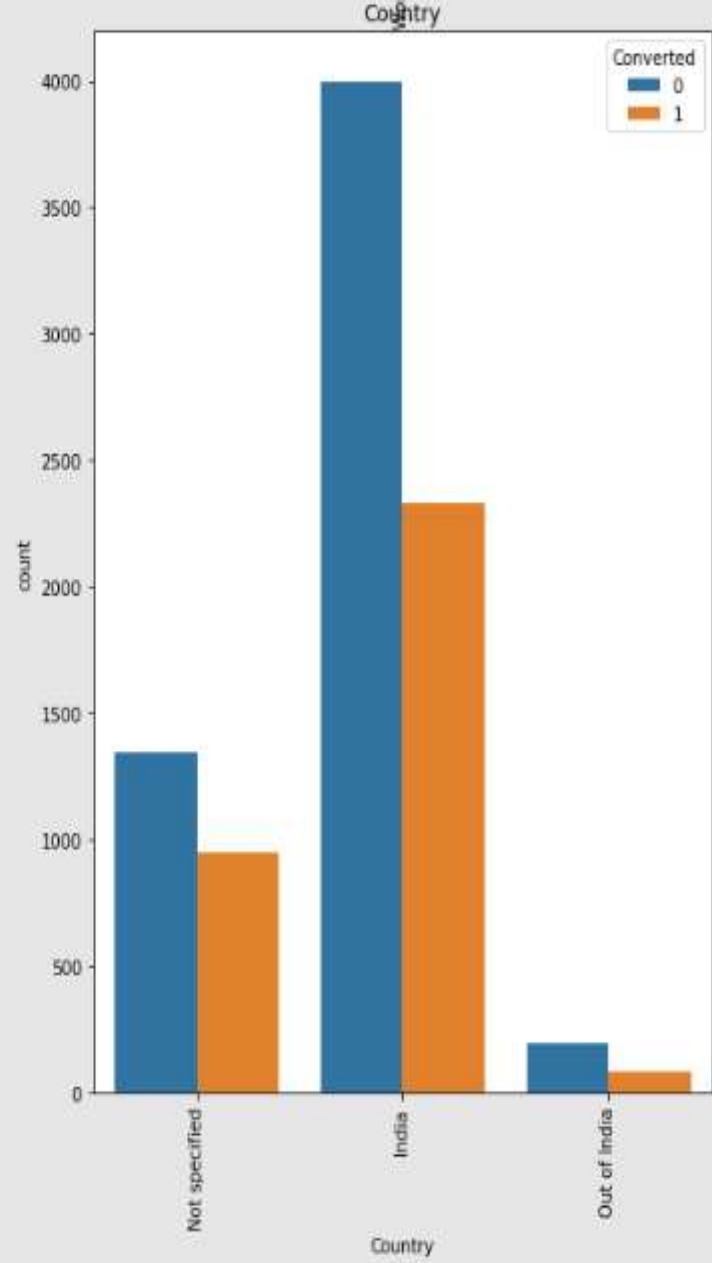
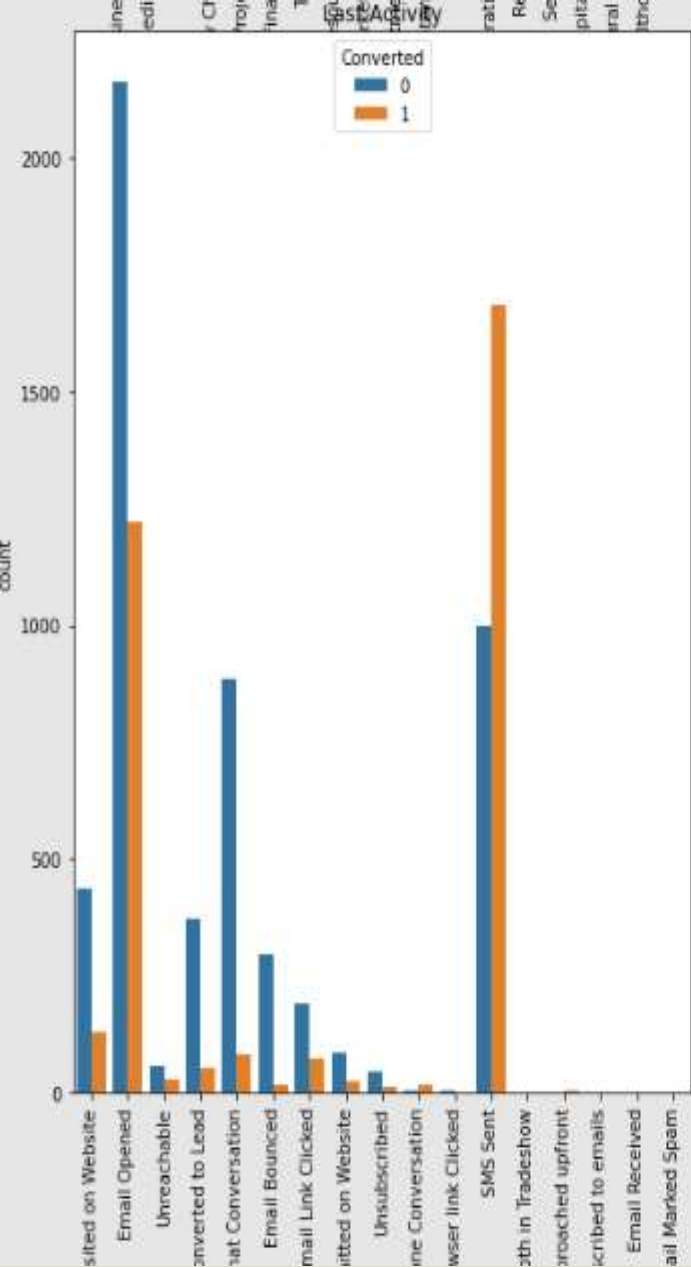


### Specialization Vs Converted:

Most of leads are not specified category and marketing management, human resources management

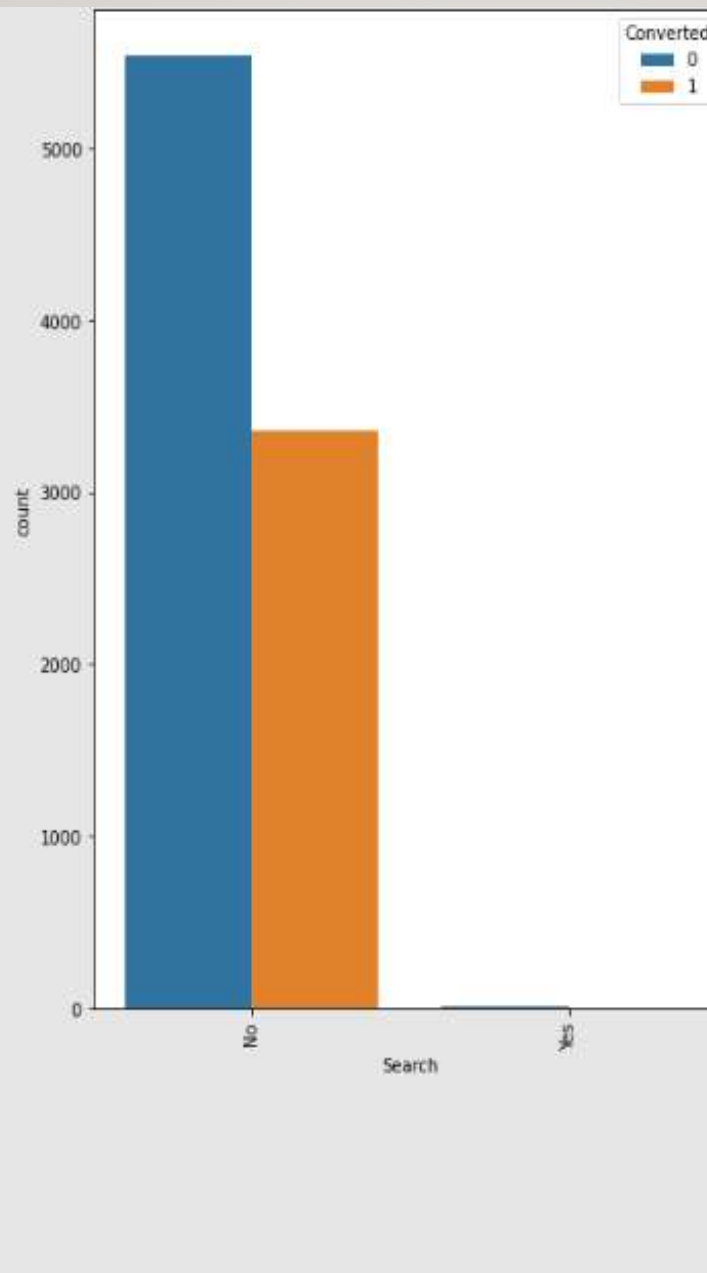
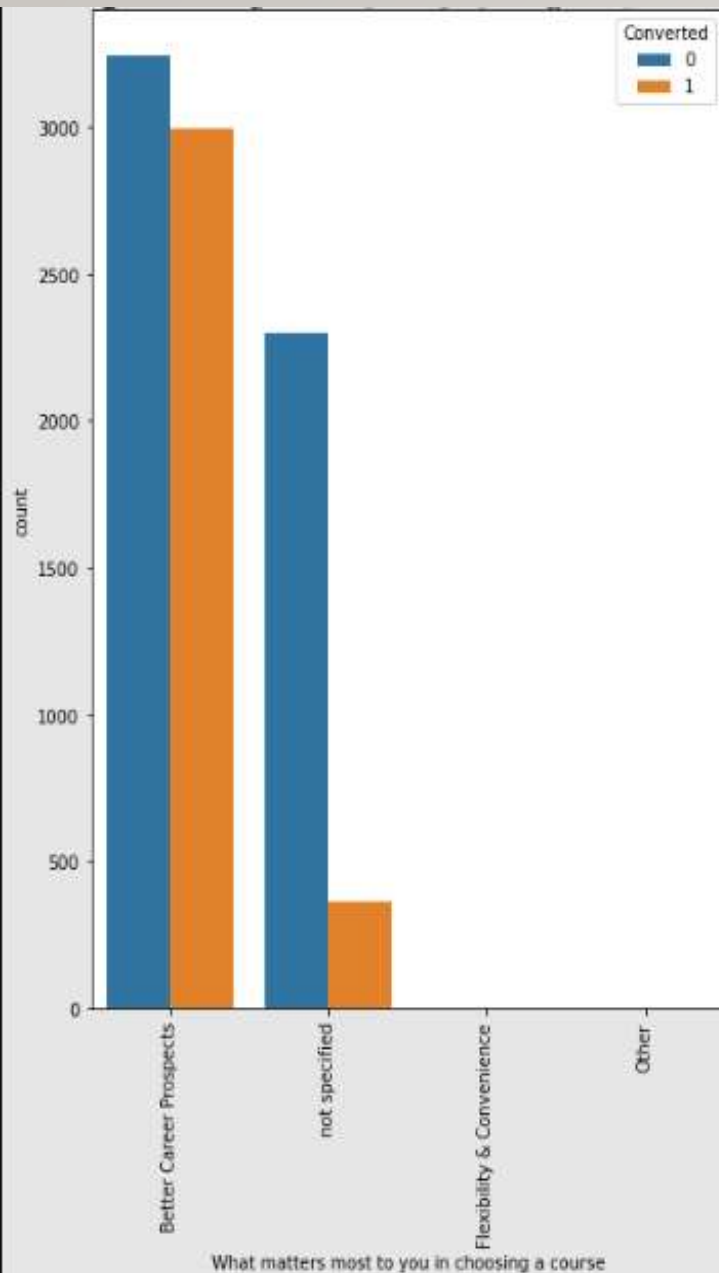
### Occupation Vs Converted :

Most of the Leads are converted by Unemployed category



Last activity VS Converted:  
SMS Sent have good Lead converted number

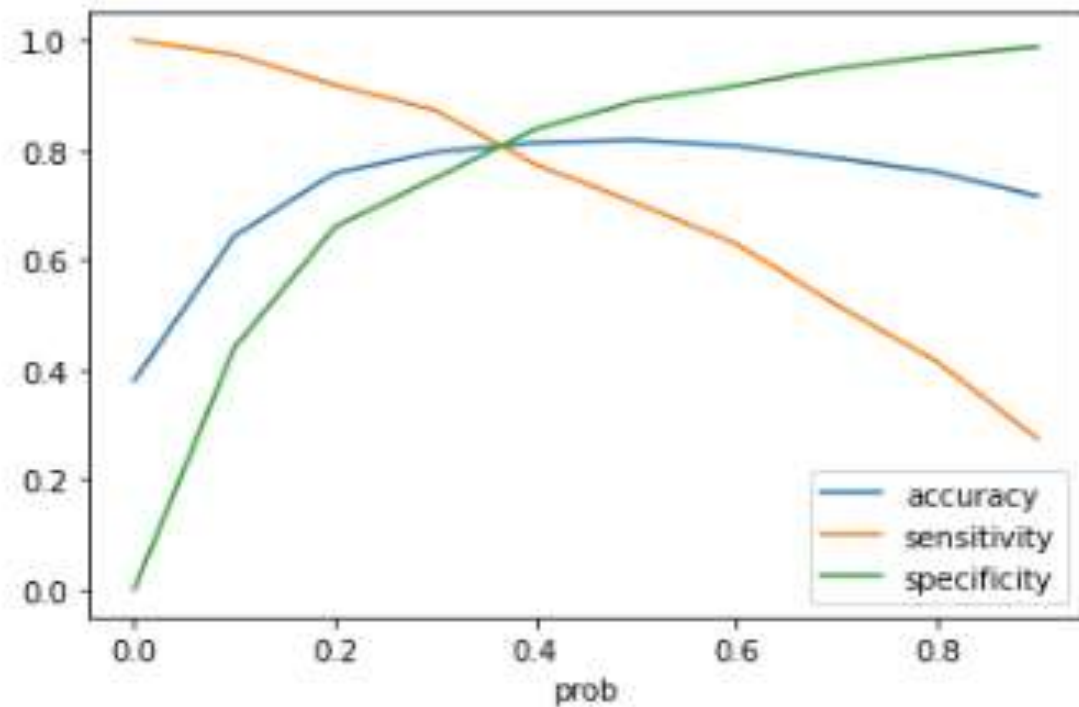
Country vs Converted:  
India has High lead conversions



What matters most to you choosing course  
Vs Converted:  
Better Career Prospects have high converted leads

Search VS Converted:  
Searches are not good enough for lead conversion rate

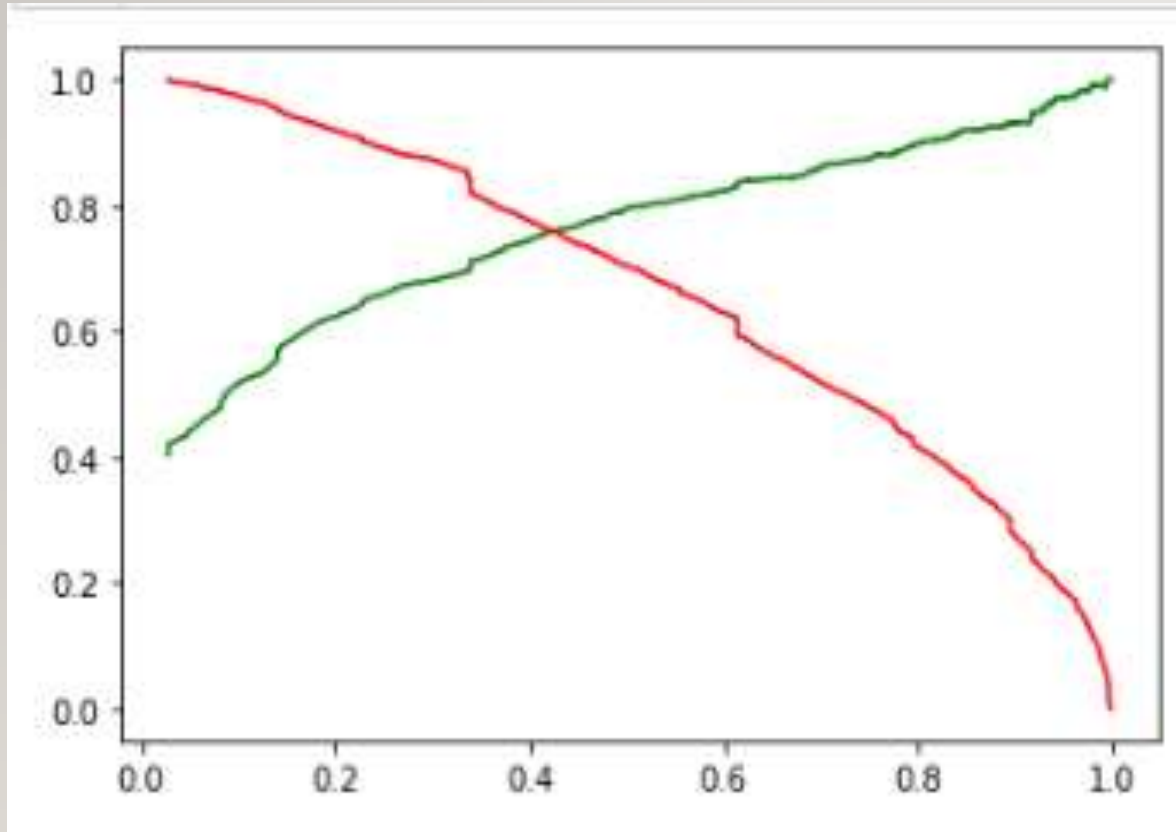
## Model Evaluation sensitivity and Specificity on Train Data:



With cutoff of 0.3 value

- ❑ Accuracy = 79%
- ❑ Sensitivity = 87%
- ❑ Specificity = 74%

## Precision and Recall:



□ Precision : 68%

□ Recall : 87 %



## Predictions on Test data :

- ❑ Accuracy : 81%
- ❑ Sensitivity is 86%
- ❑ Specificity is 78%



# Conclusion:

## EDA Analysis:

Landing pages can be helpful to find more leads  
SMS Messages have high impact on the leads conversions  
Searches are not enough for lead conversions  
India has good lead conversion rate

## Model Analysis:

Model Shows accuracy of 79.6%  
Using measures, the threshold is opted from sensitivity ,specificity ,precision and recall  
Model has 87% sensitivity and 74% Specificity