# Lead Scoring Case Study

Ashutosh Kumar

Introduction

- An education company named "X Education" sells **online courses** to industry professionals

- The company gets its leads through **past referrals** or websites and **search engines**

- 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%**

- The CEO, in particular, has given a **ballpark** of the target lead conversion rate to be around **80%**

- We will be using **logistic regression** to build the model by using **"Converted"**as the **target** variable and the remaining variables as the independent variables

- We have used **WOE** approach to encode categorical variables

- The **model performance** will be judged using metrics such as sensitivity, recall, precision

Important Variables

Following **variables** contribute the most towards the **probability** of a lead getting **converted**:
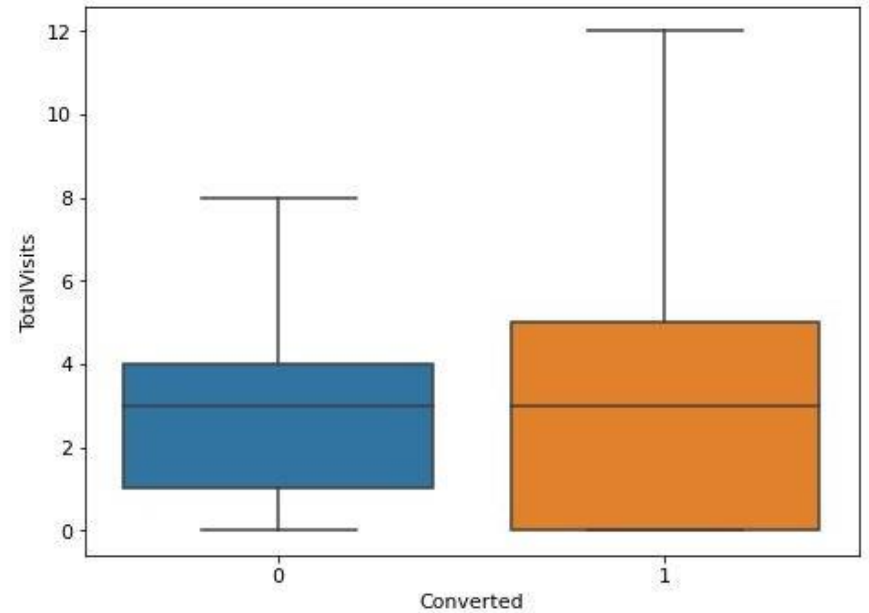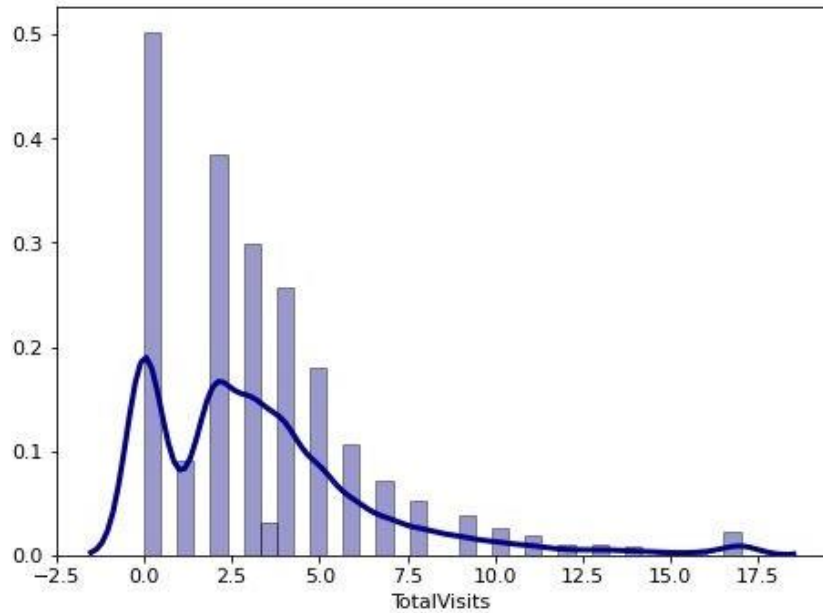
- Total Time Spent on Website
- Current occupation
- Do Not Email
- Last Activity
- Lead Source
- Page Views Per Visit
- Last Notable Activity
- Total Visits
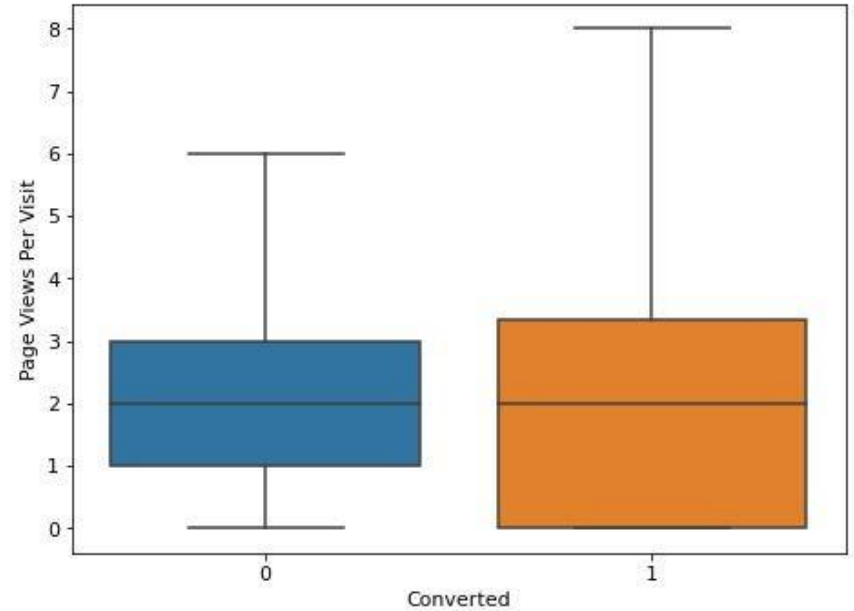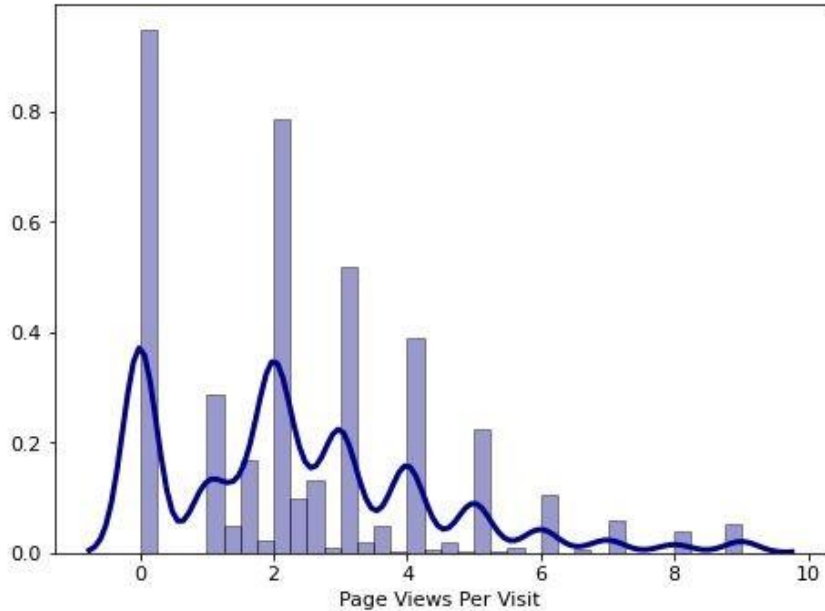- Specialization

Total Time Spent on Website
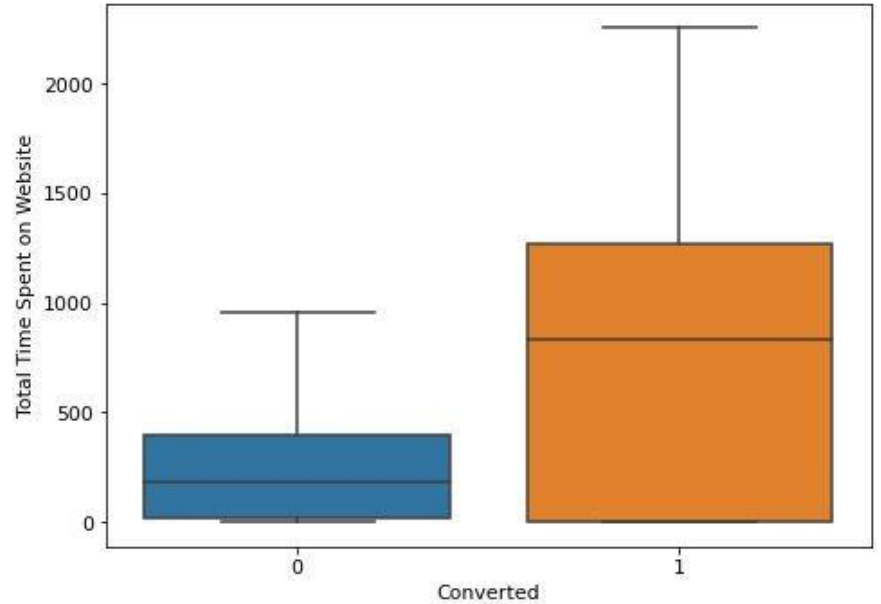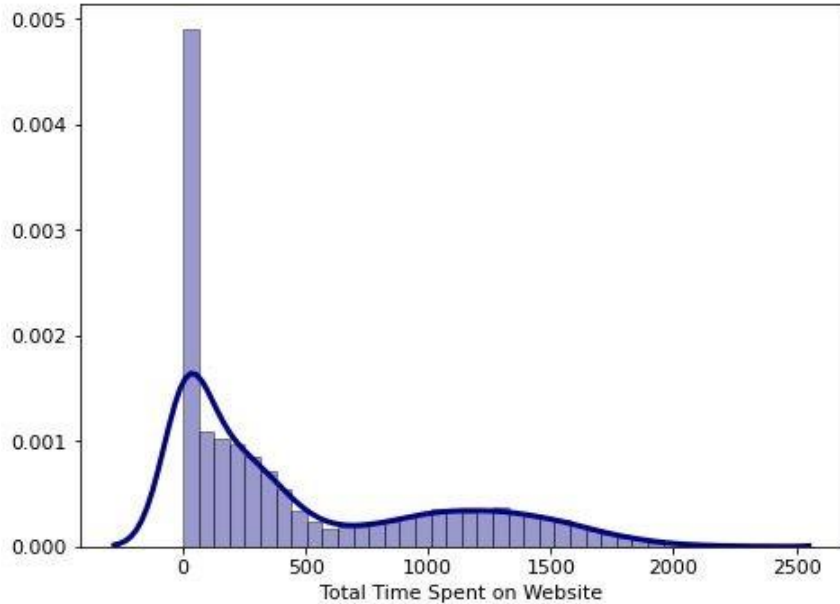
The candidates who spend more time on the website show higher rate of conversion
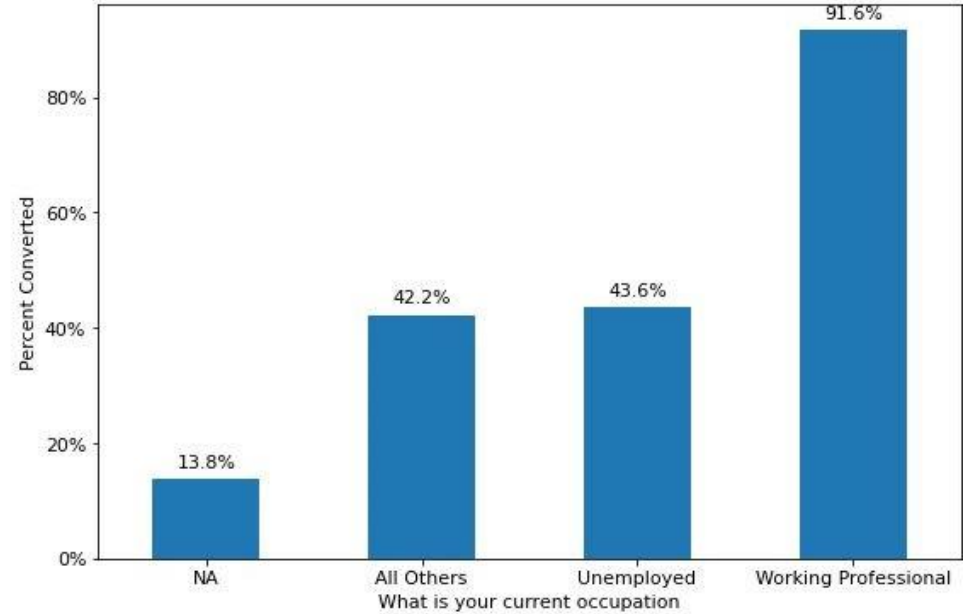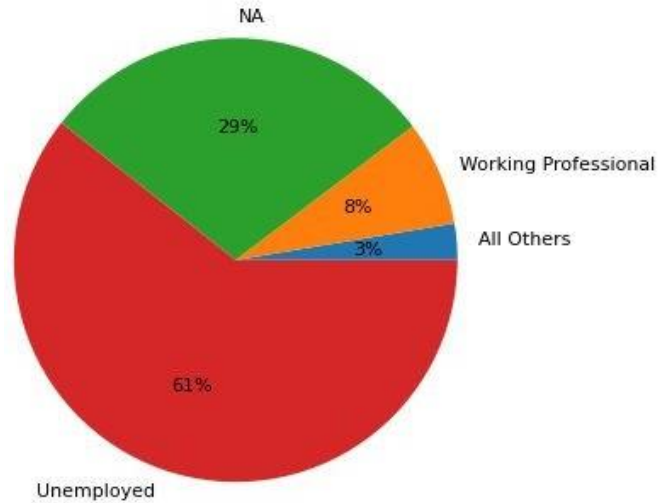
Total Visits

The candidates who visit the website more often show slightly higher rate of conversion
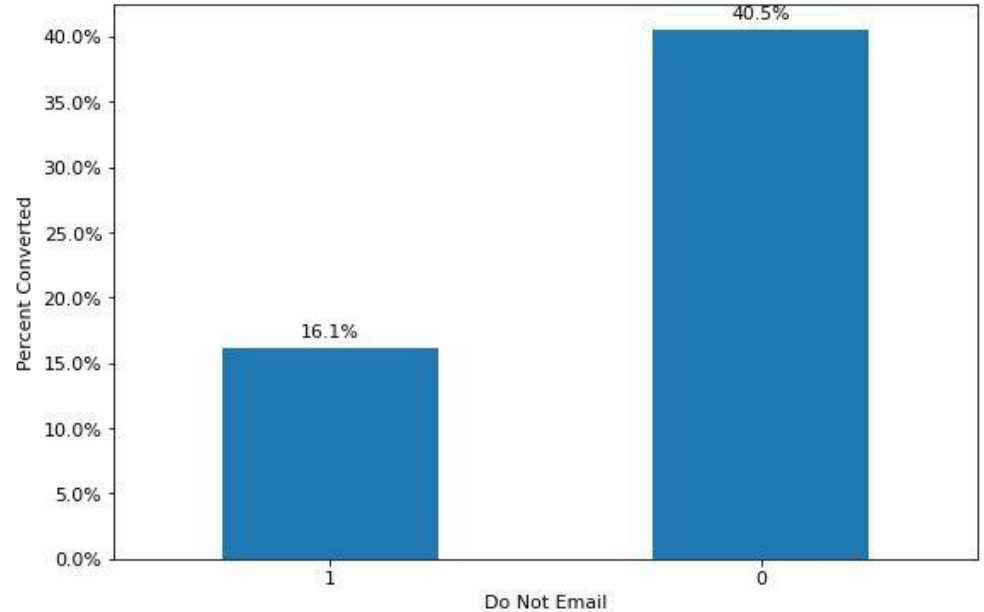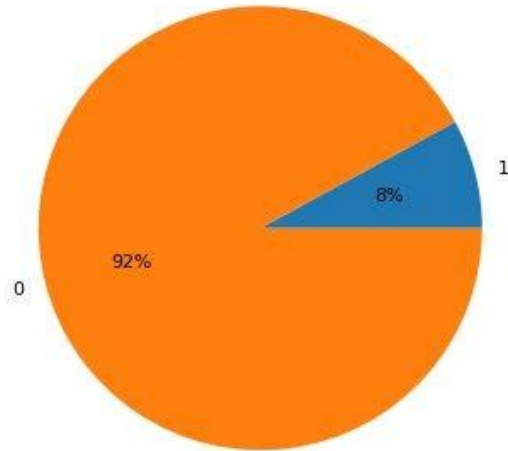
Page Views Per Visit

The candidates who view more website pages per visit show slightly higher rate of conversion

Current Occupation

The candidates who are **working professionals** have the **highest** rate of conversion whereas those whose occupation is **NA (Not Available)** have the **lowest** rate of conversion
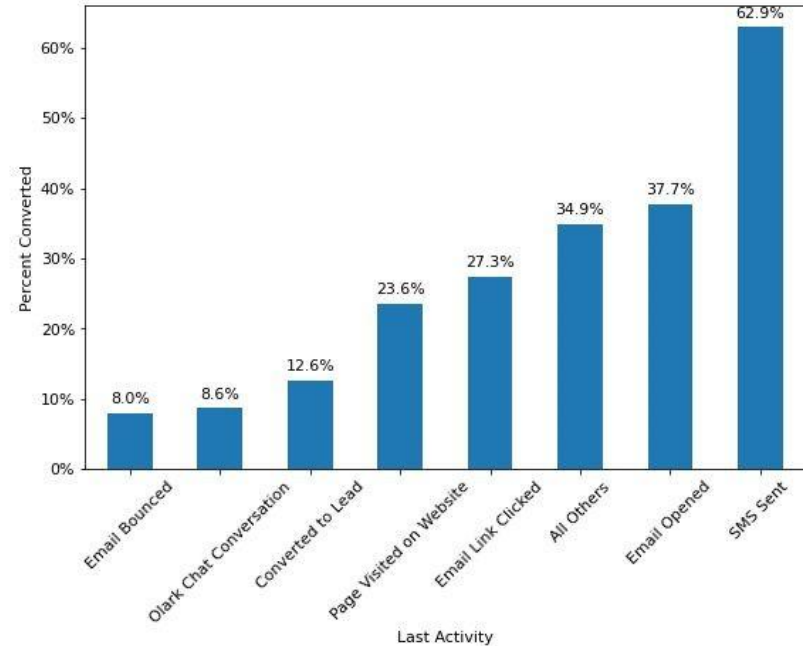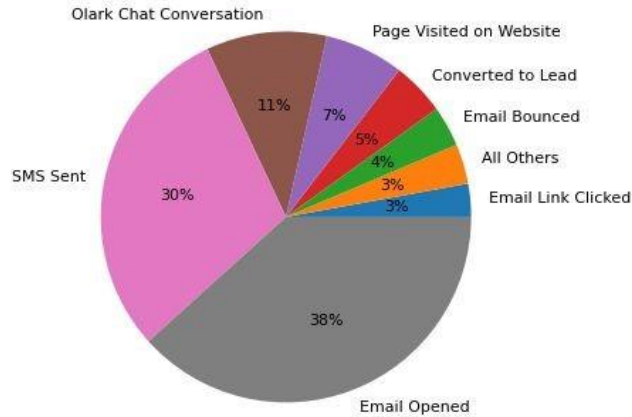
Do Not Email

The candidates who opt **to receive** emails about the course have the **highest** rate of conversion whereas those who opt **not to receive** emails have the **lowest** rate of conversion
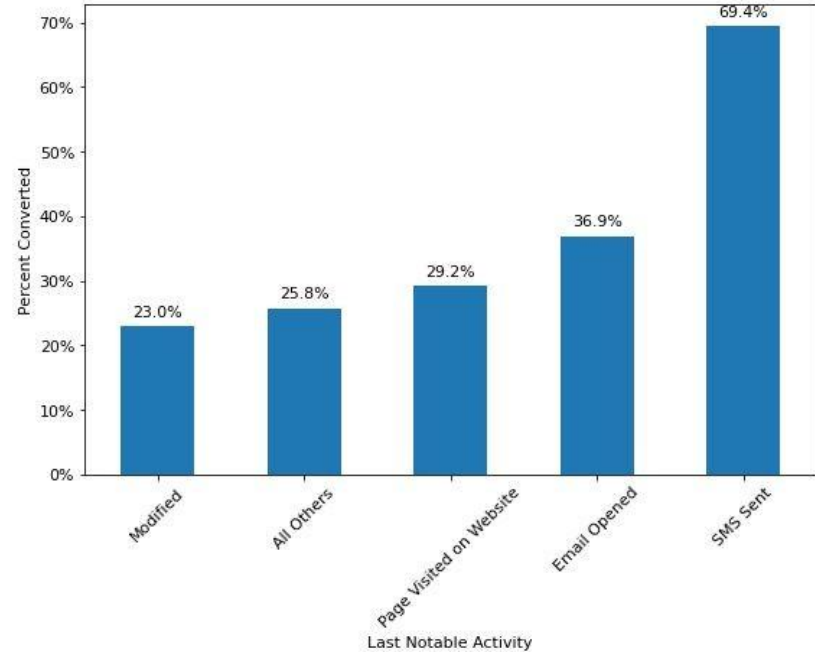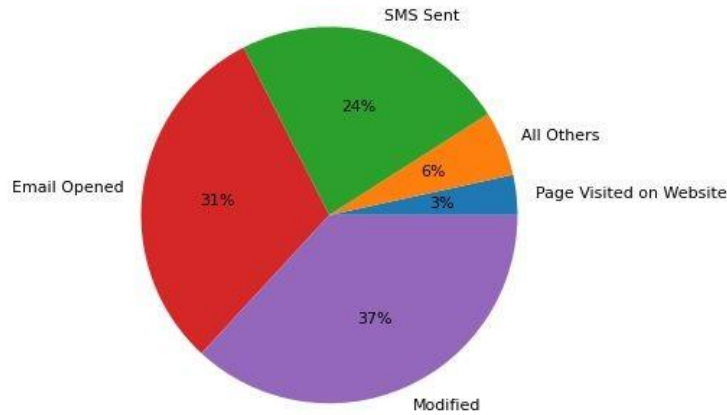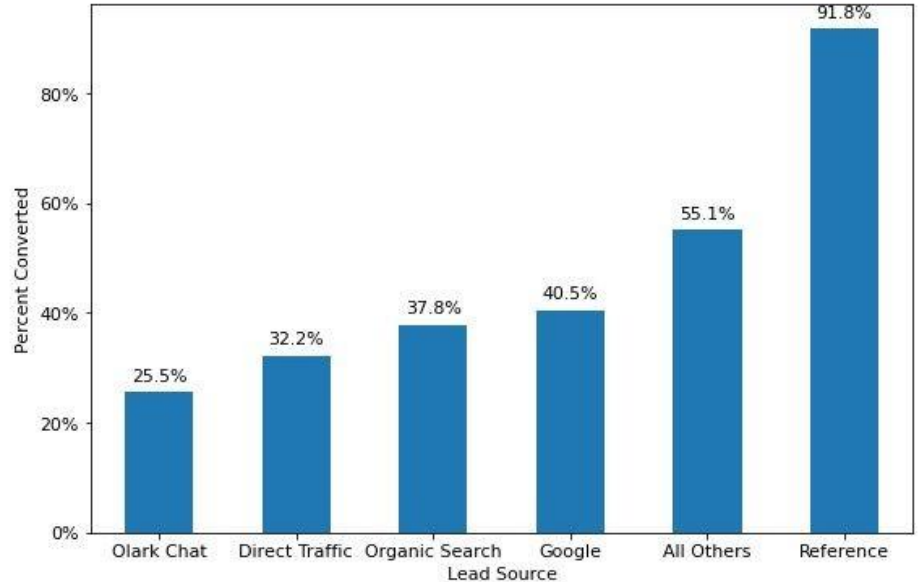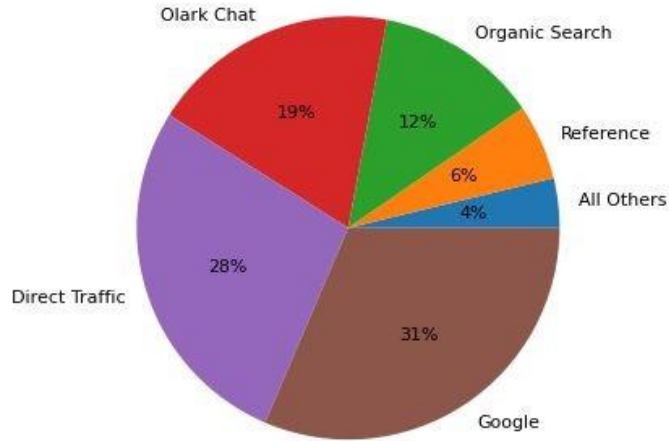
LastActivity

The candidates whose last activity is "SMS Sent" have the **highest** rate of conversion whereas those whose last activity is "Email Bounced" have the **lowest** rate of conversion
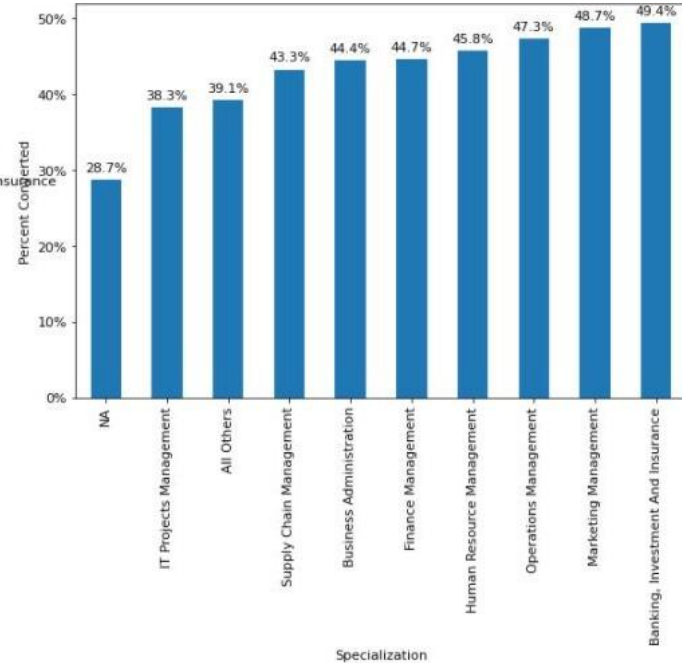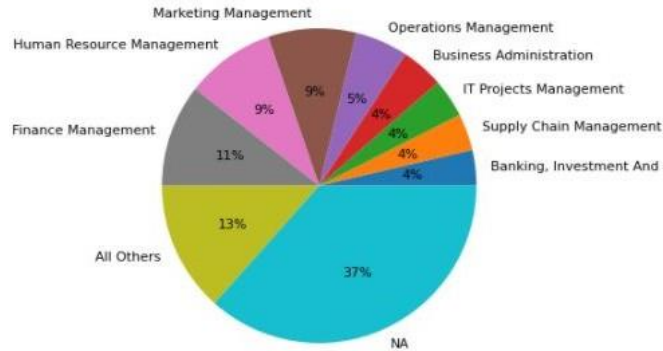
Last Notable Activity

The candidates whose Last Notable Activity is "SMS Sent" have the **highest** rate of conversion whereas those whose Last Notable Activity is "Modified" have the **lowest** rate of conversion

Lead Source

The candidates whose lead source is "Reference" have the **highest** rate of conversion whereas those whose lead source is "Olark Chat" have the **lowest** rate of conversion
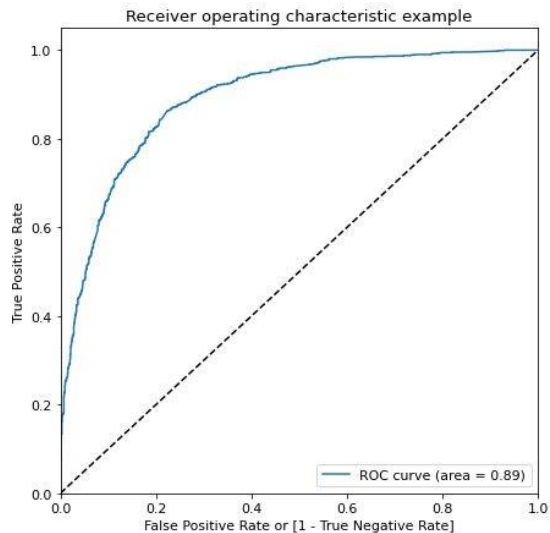
Specialization

The candidates whose Specialization is "Banking,Investment and Insurance" have the **highest** rate of conversion whereas those whose Last NotableActivity is "Modified" have the **lowest** rate of conversion
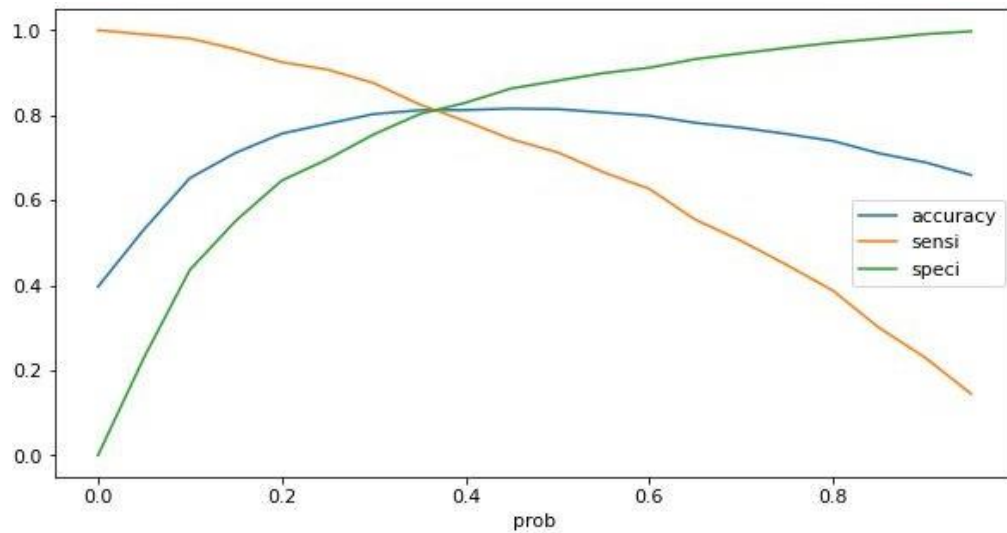
Country
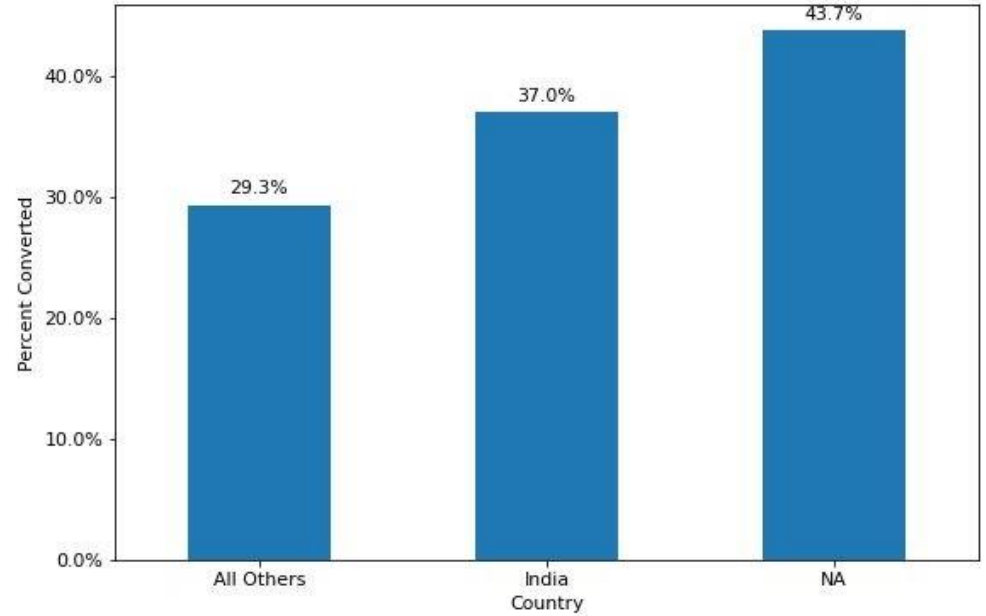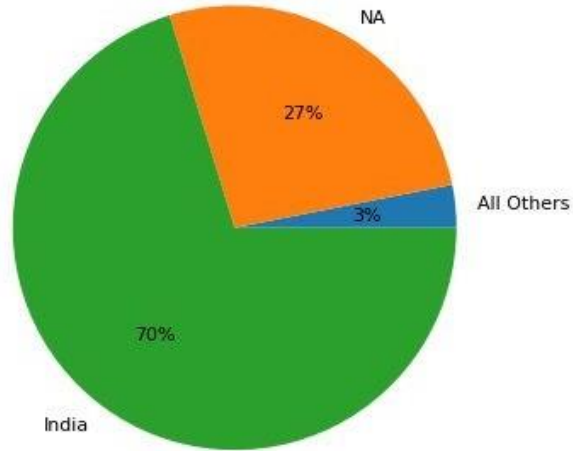
The candidates from countries other than India have the **lowest** rate of conversion

ROCAUC

The ROC AUC of the model is 0.89

The optimal cut-off is at 0.35. The accuracy, sensitivity and specificity at this cut-off are approximately 0.8

Recommendations

We recommend the sales team to focus on candidates who:

- Are more engaged on the website (spend more time, visit more often, visit more pages per visit)
- Are working professionals
- Opt to receive emails about the course
- Have their last activity as "SMS Sent"
- Came through a reference

The ideal approach would be to start calling prospects with the highest predicted conversion probability first and go down that list as far as required.