LEAD SCORING CASE STUDY

Nikhil Mahakal

AGENDA

INTRODUCTION
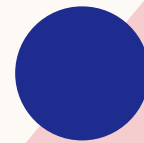
- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process more efficient, the company wishes to identify the 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.

PRIMARY GOALS

- X education wants to know most promising leads.
- For that they want to build a Model which identifies the hot leads.
- Deployment of the model for the future use.

WORKING METHODOLOGY

- Data Cleaning and Data Manipulation
  - ❖ Check and handle duplicate data.
  - ❖ Check and handle missing values.
  - ❖ Drop unnecessary columns that contain large number of missing values and is useless for Analysis.
  - ❖ Imputation of values if required.
  - ❖ Check and handle Outliers in data.
- Exploratory Data Analysis
  - ❖ Univariate Data Analysis : value count, distribution of variable.
  - ❖ Bivariate Data Analysis : correlation coefficient, pattern between the variable.
- Classification Technique : Logistic Regression used for model building and prediction.
- Validation of Models.
- Model Presentation.

EXPLORATORY DATA ANALYSIS
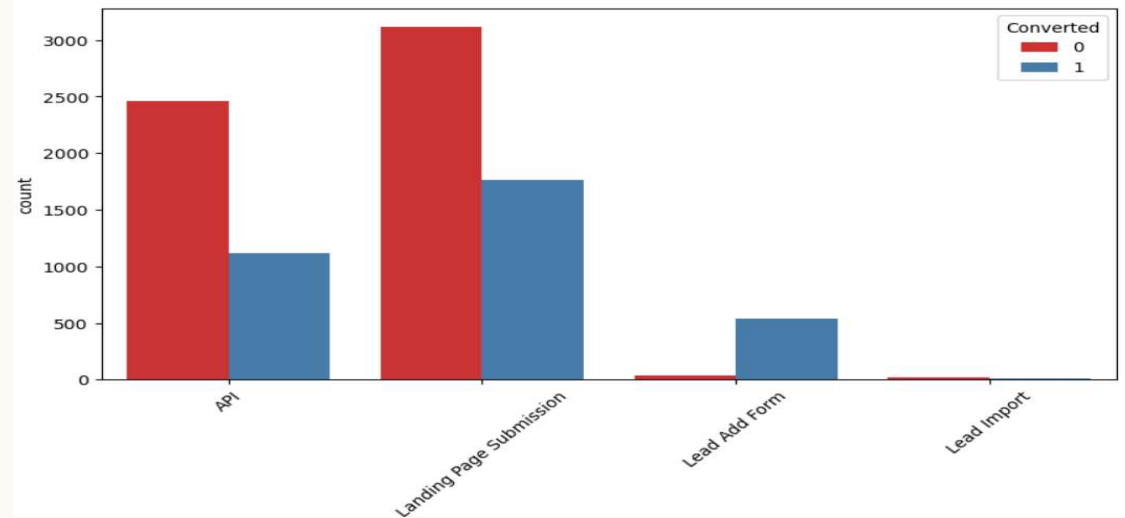
Inference :

- API and Landing Page Submission have 30-35% conversion rate but count of lead originated from them are considerable.
- Lead Add Form has more than 90% conversion rate but count of lead are not very high.
- Lead Import are very less in count.

To improve overall lead conversion rate, we need to focus more on improving lead converion of API and Landing Page Submission origin and generate more leads from Lead Add Form.
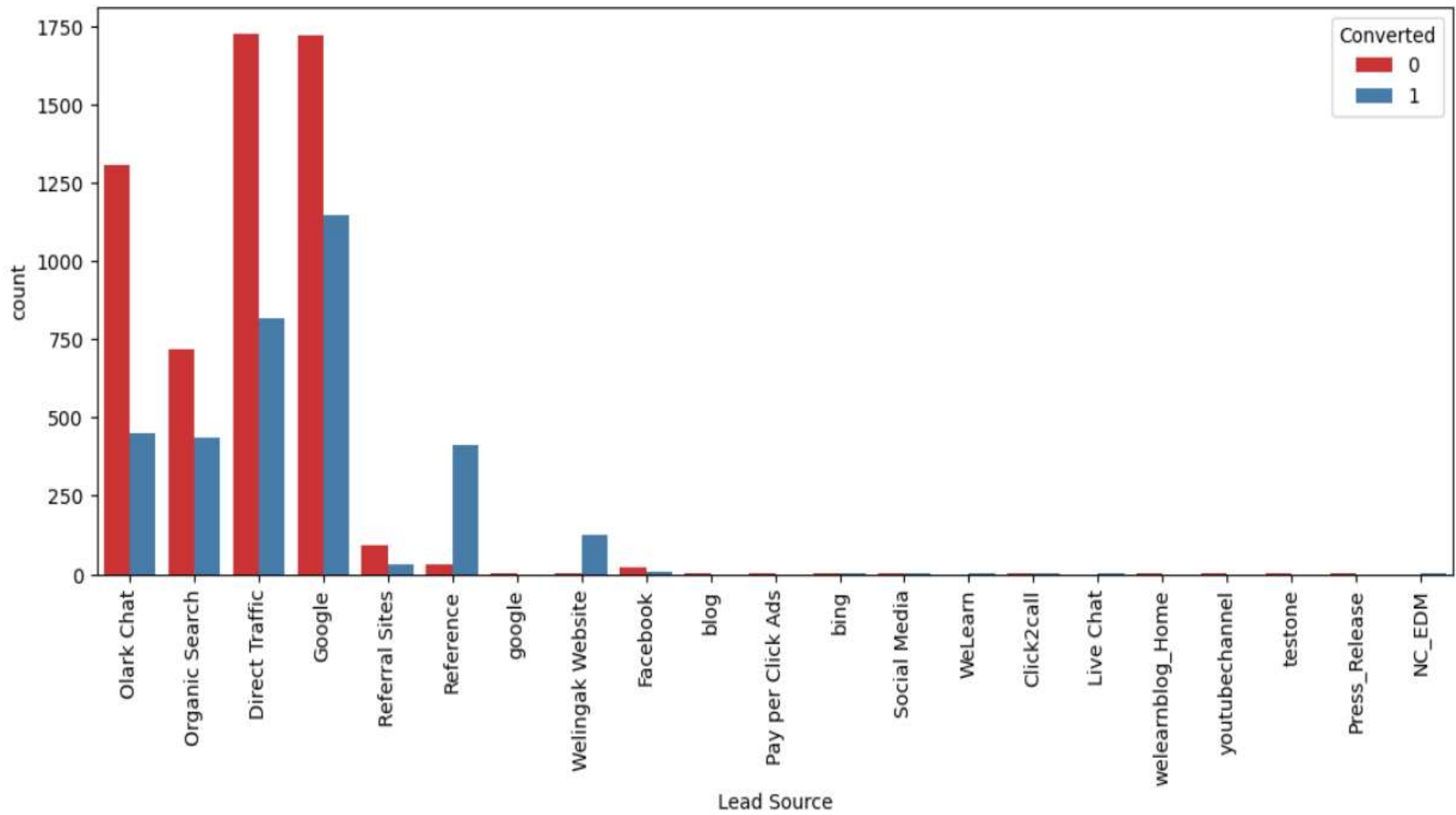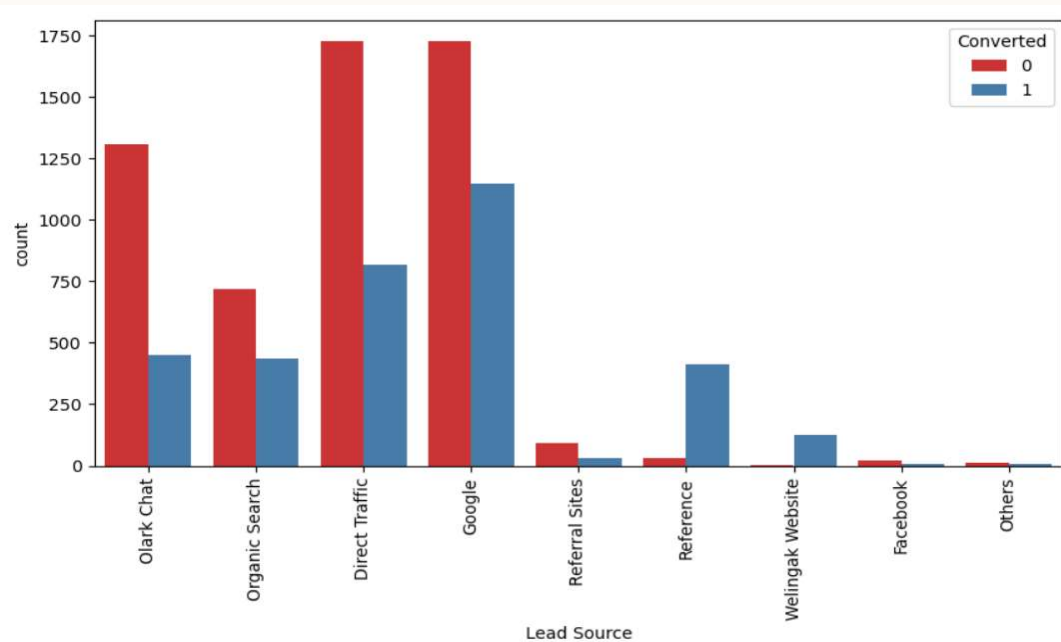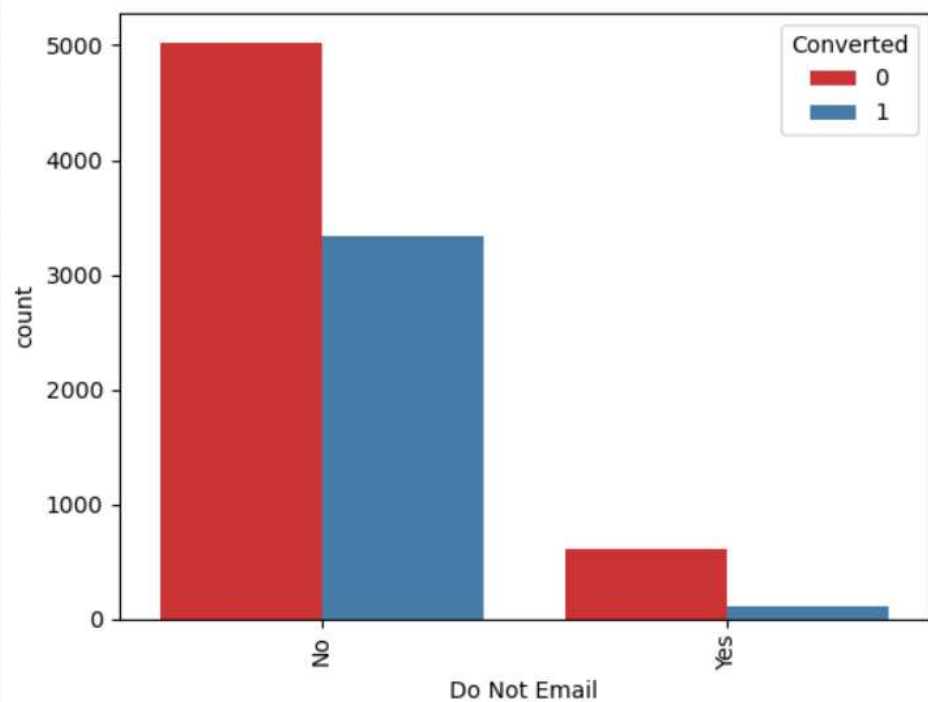
Inference

- Google and Direct traffic generates maximum number of leads.
- Conversion Rate of reference leads and leads through welingak website is high.

To improve overall lead conversion rate, focus should be on improving lead converion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.
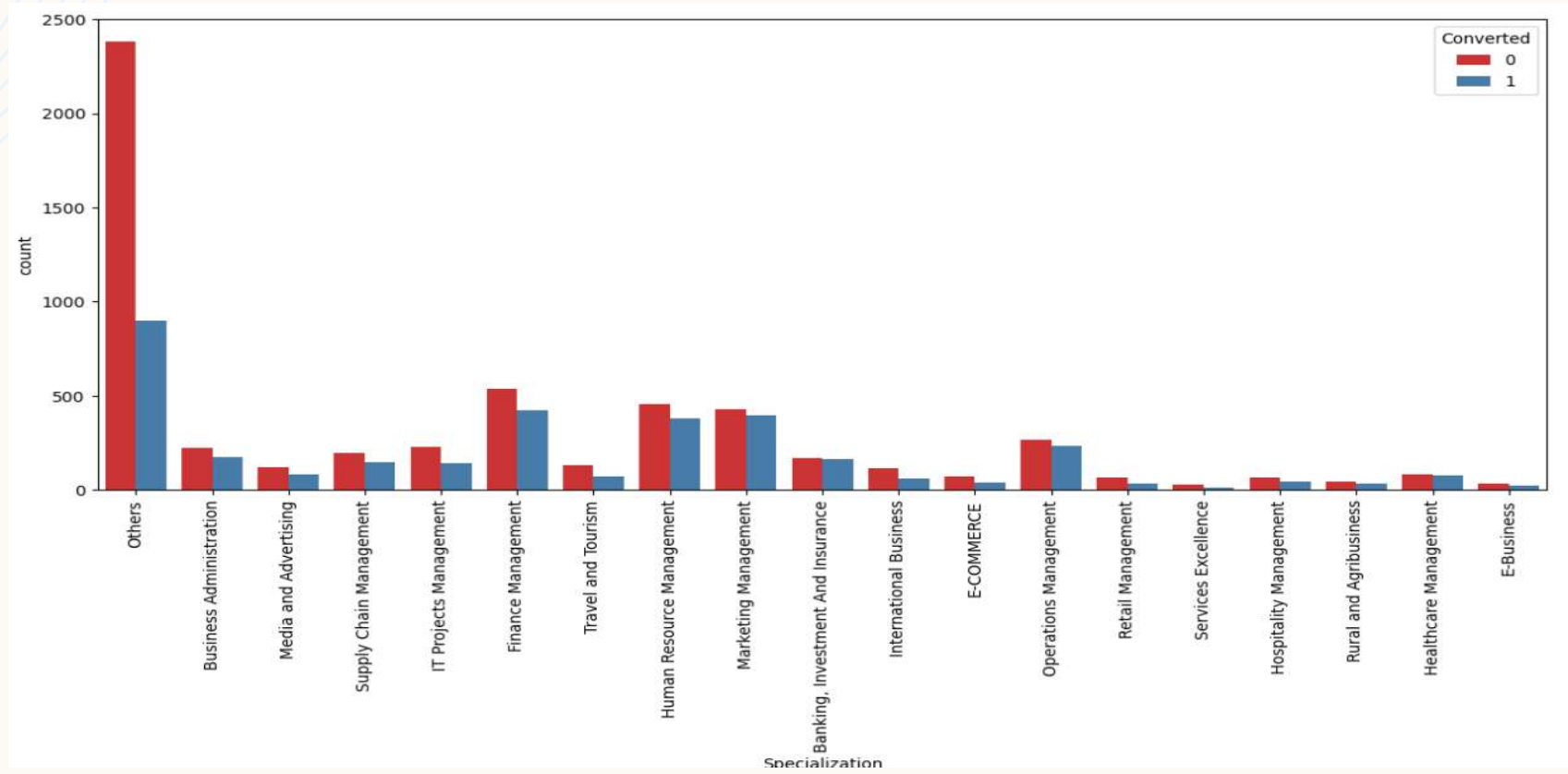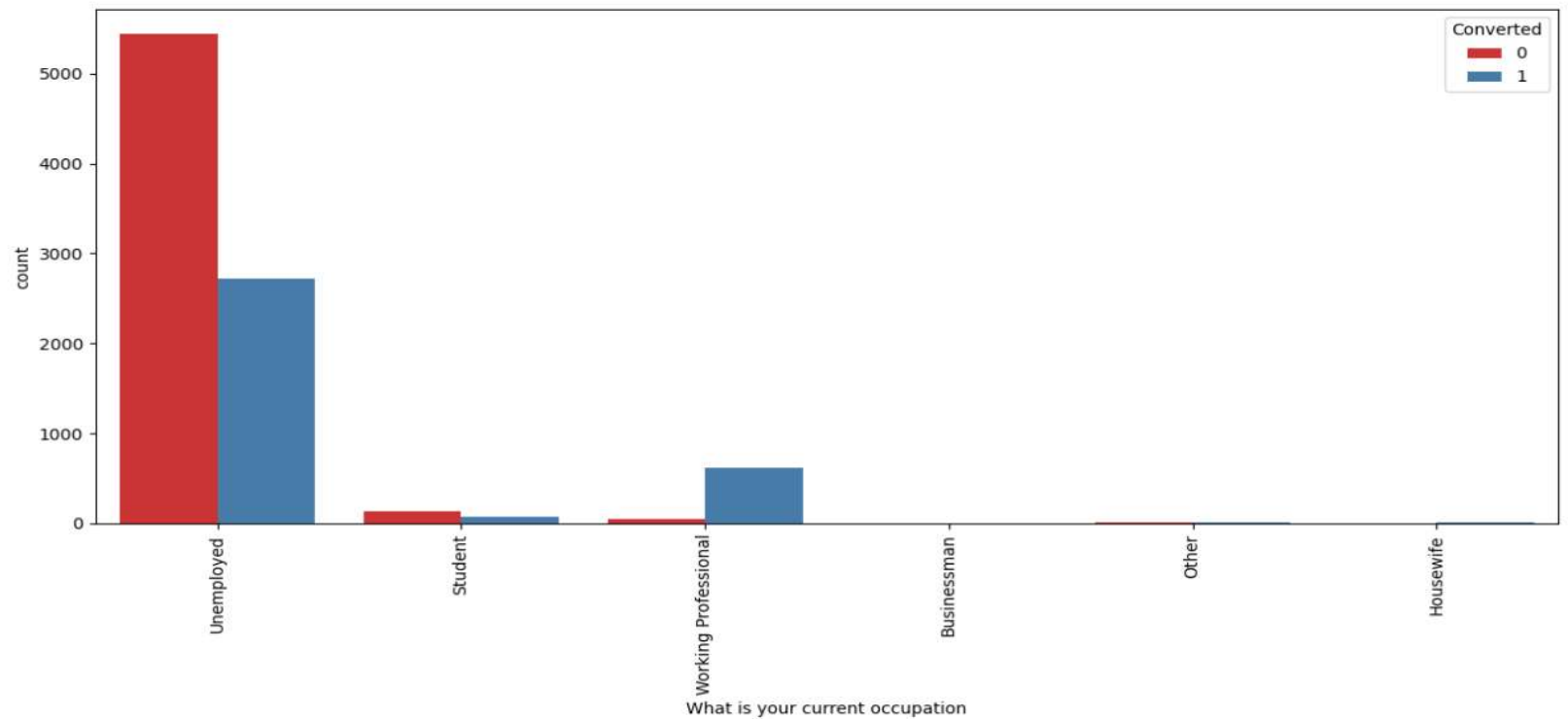
Inference

- Most entries are 'No'. No Inference can be drawn with this parameter.
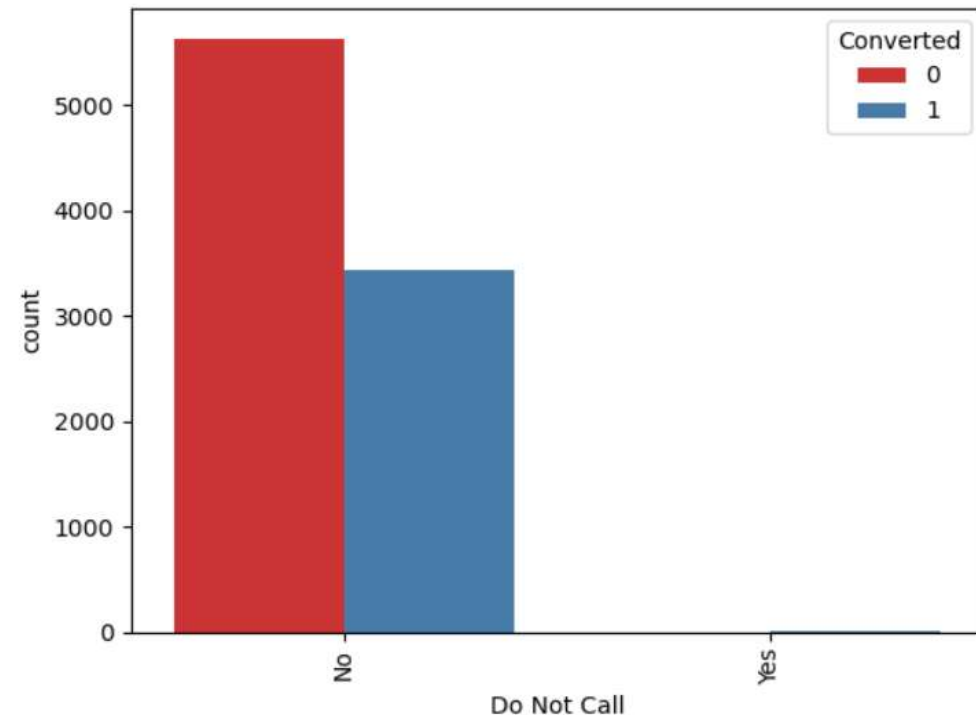
Inference

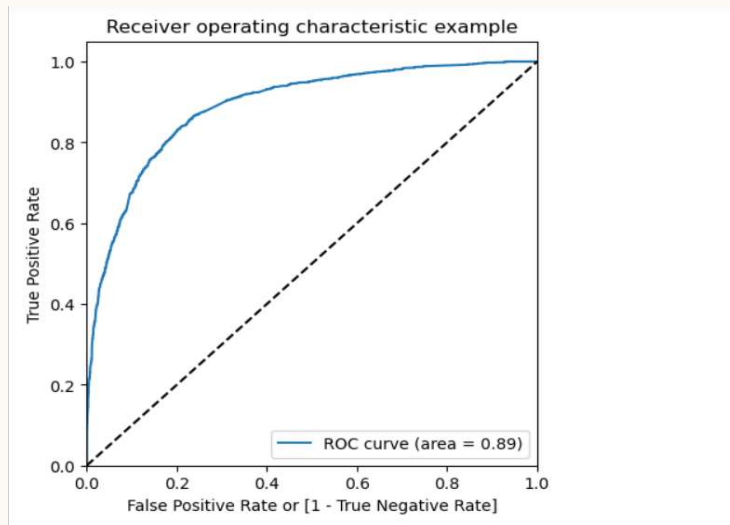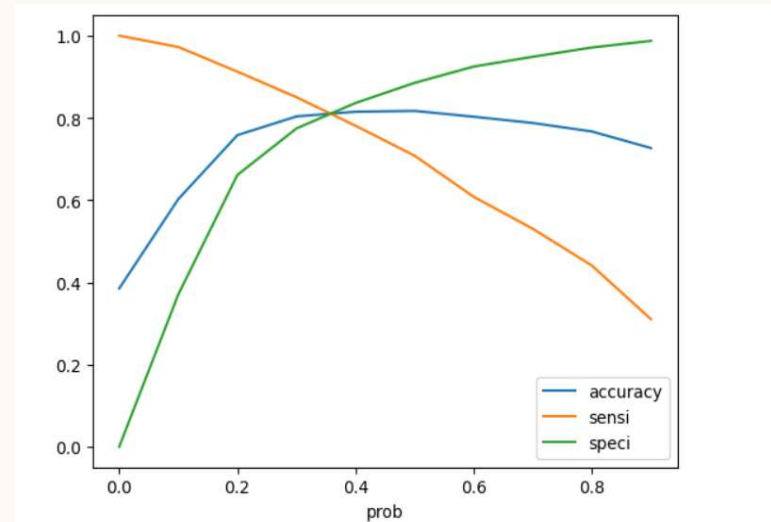- Most entries are 'No'. No Inference can be drawn with this parameter.

Inference

- Working Professionals going for the course have high chances of joining it.
- Unemployed leads are the most in numbers but has around 30-35% conversion rate.

ROC CURVE

Since we have higher (0.89) area under the ROC curve , therefore our model is a good one.

From the curve above, 0.34 is the optimum point to take it as a cutoff probability.

- Optimal cut off probability is the point where we get balanced sensitivity and specificity.
- Here the Optimal cut off is 0.35.

CONCLUSION

- We can conclude that the variables that influence most of the potential buyers are:

- Total time spent on website.

- Total number of visits.

- When last activity was SMS, Olark Chat Conversation.

- When the lead source where Google, Direct Traffic, Organic Search, Welingak Website.

- When their current occupation is as a working professional. Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.