LEAD SCORING CASE STUDY

PRESENTED BY –

MR. SHIKHAR JAIN

MR. OMKAR LONDHE

• Problem Statement –

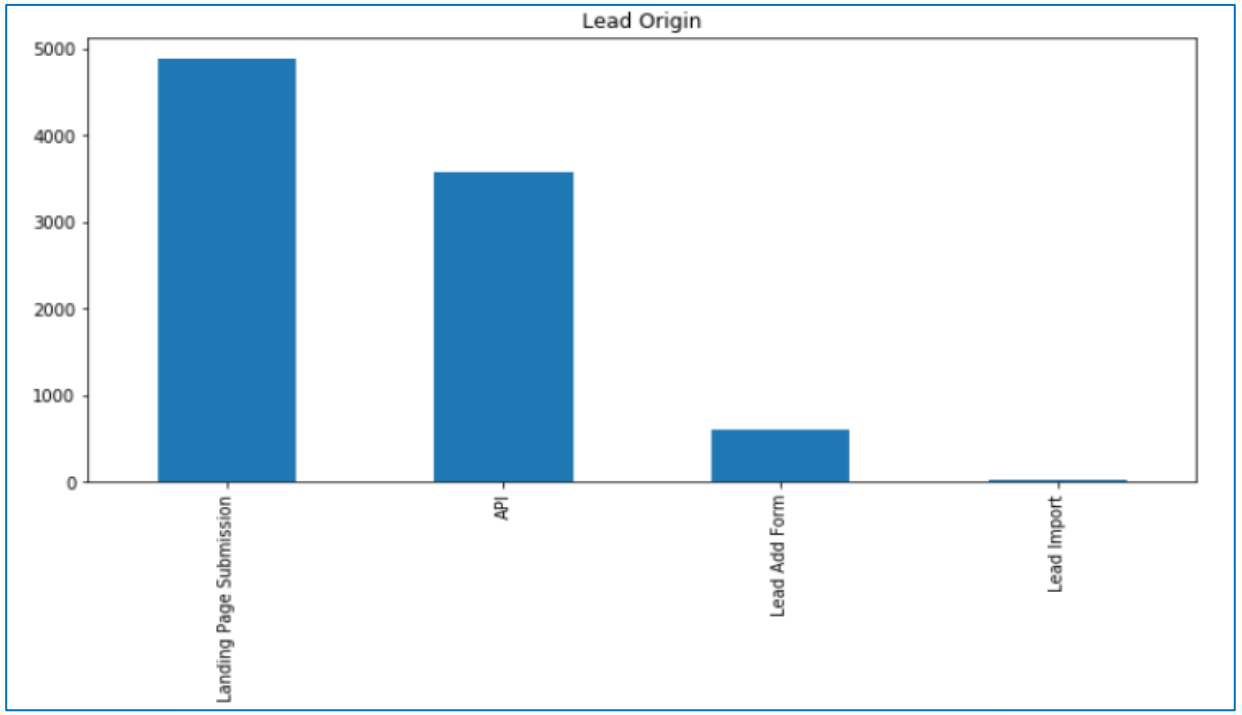
- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- The company markets its courses on several websites and search engines like Google to generate the leads. 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% & Company wants to increase it to 80%.

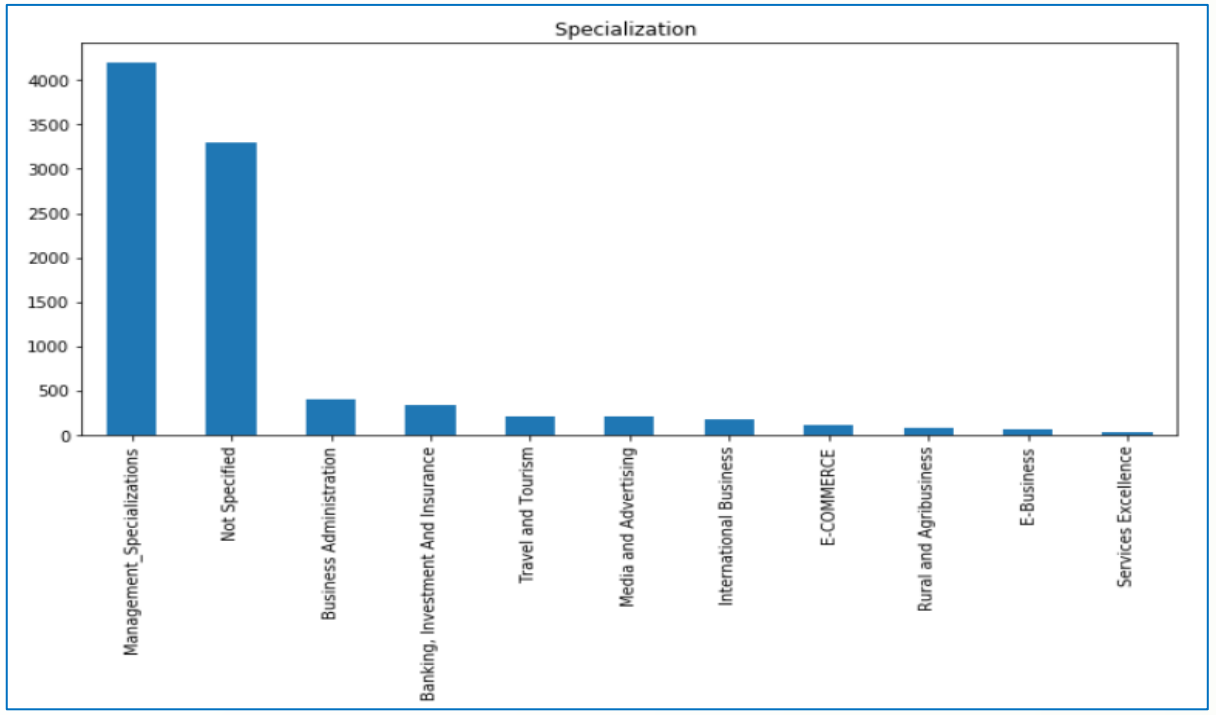
• Objective –

- 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.

- **EDA –**

Univariate Analysis =



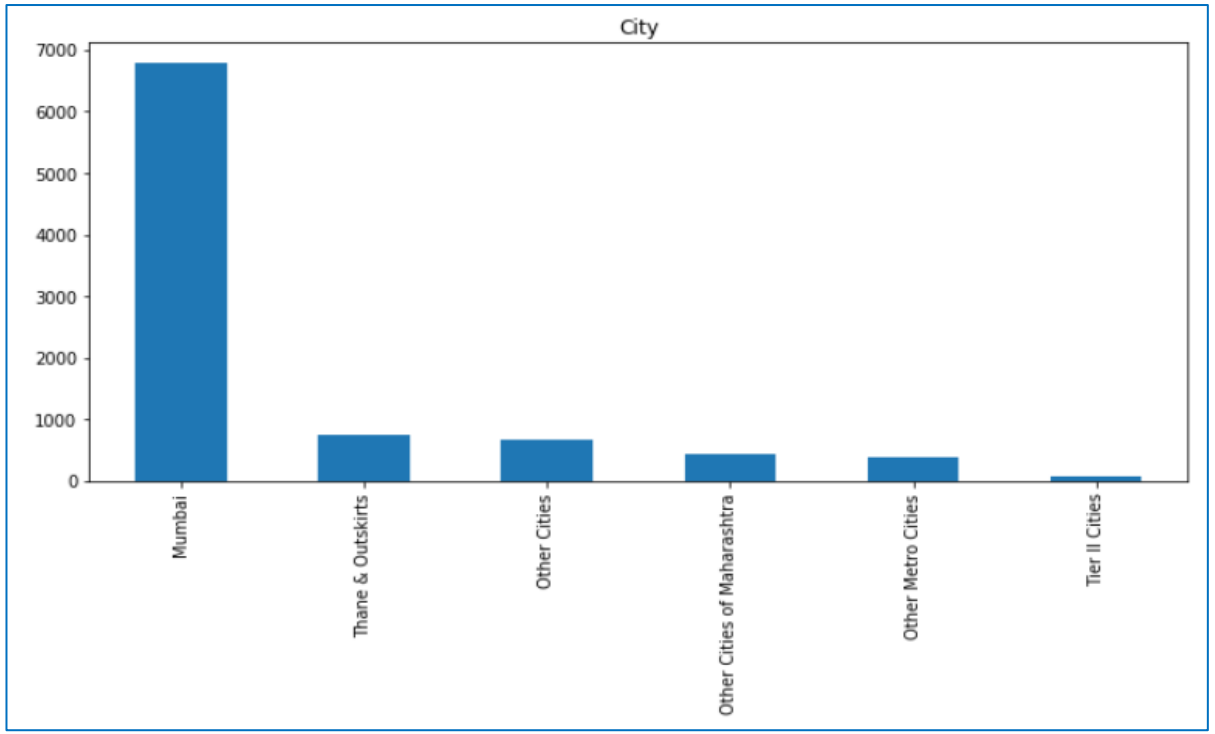


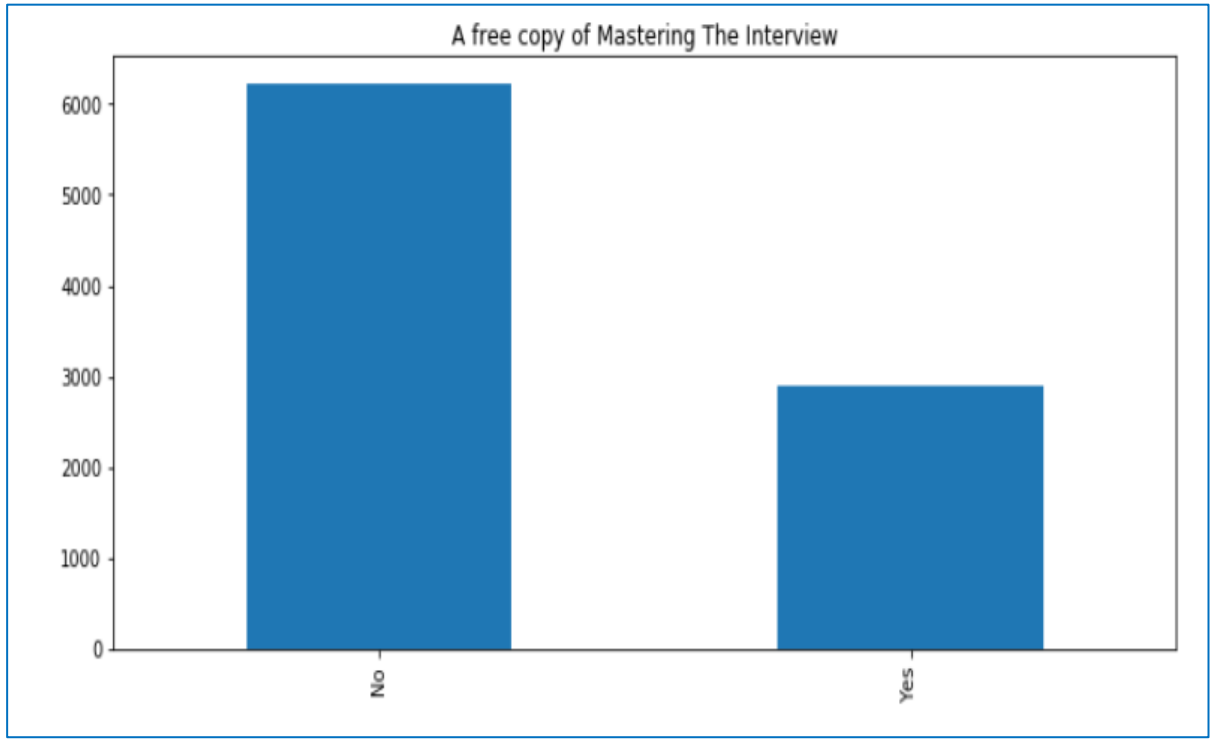
- Most No of leads are generated from Landing Page Submission & API

- Large no of customer have specialization of management.
- Also many customer didn't specify their specialization , this may be because these are students.

- **EDA –**

Univariate Analysis -

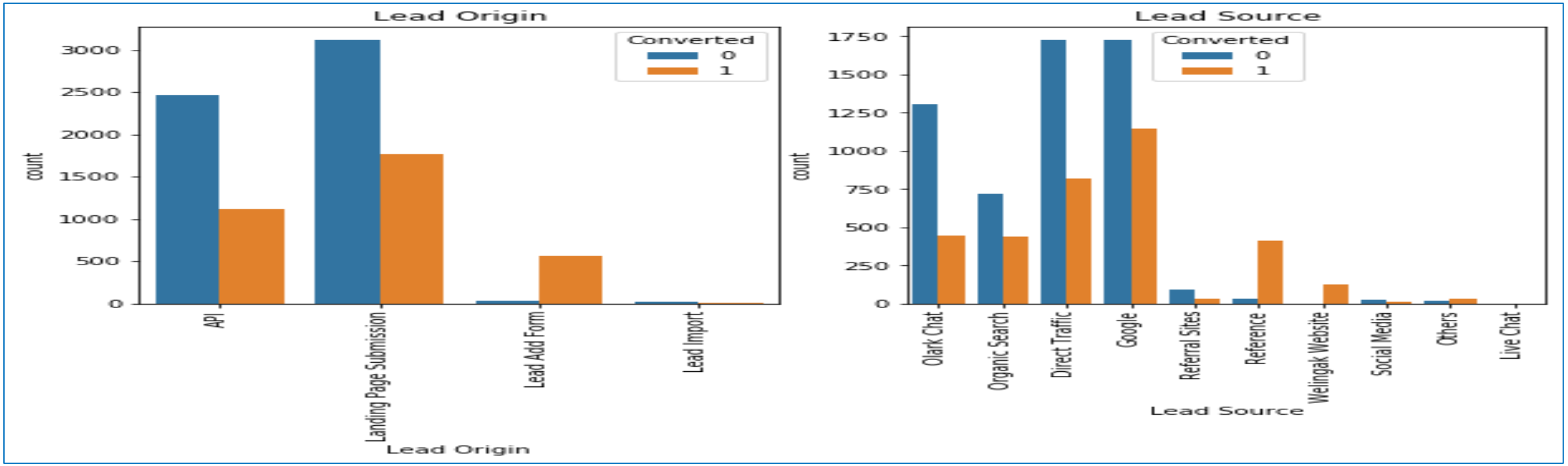




- Most No of leads are generated from Mumbai
- Inspite of being freely available , most number of customers don't want copy of Mastering The interview.

- **EDA –**

Bivariate Analysis -



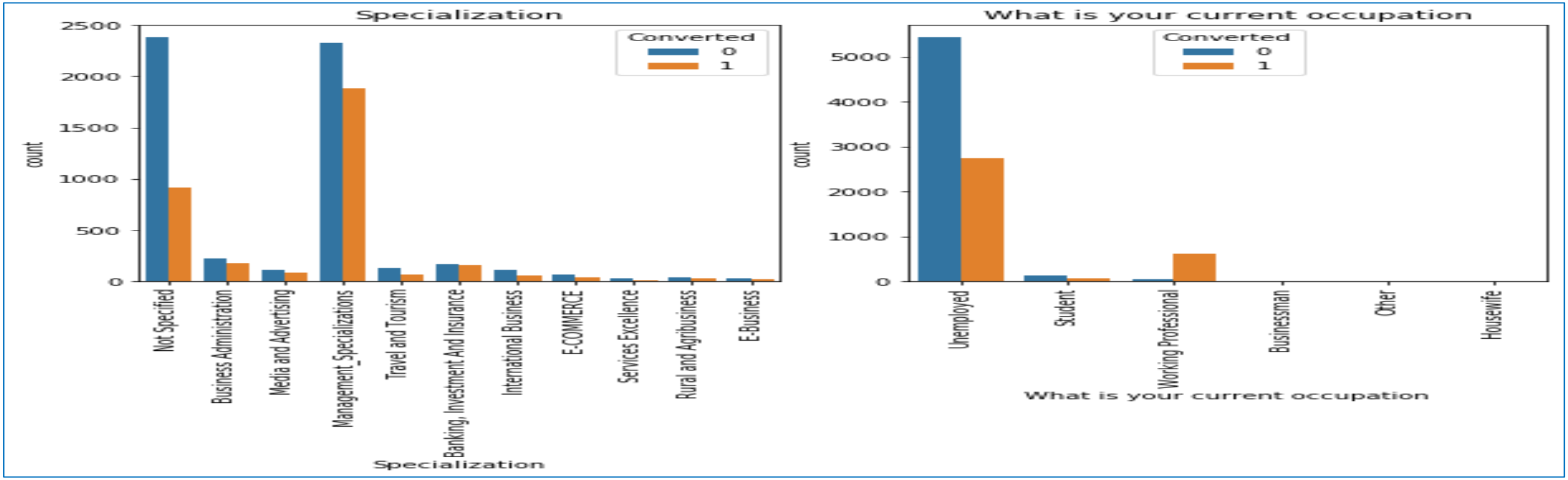
- API and Landing Page Submission have fairly good conversion rate but number of leads generated from them are higher.
- Lead Add Form has higher conversion rate than others but number of leads generated are not very high.

- To improve the Lead conversion rate , company needs to focus more on converting leads generated through API & Landing Page submission.
- Also they need to focus on generating more leads through Lead form submission as conversion rate is highest over there.

- Leads coming form souces like Google , Direct traffic , Olark Chat have good conversion rate , but number of leads generated are on higher side
- Leads coming from reference , Welingak website has higher conversion rate than others but number of leads generated are not very high.

- **EDA –**

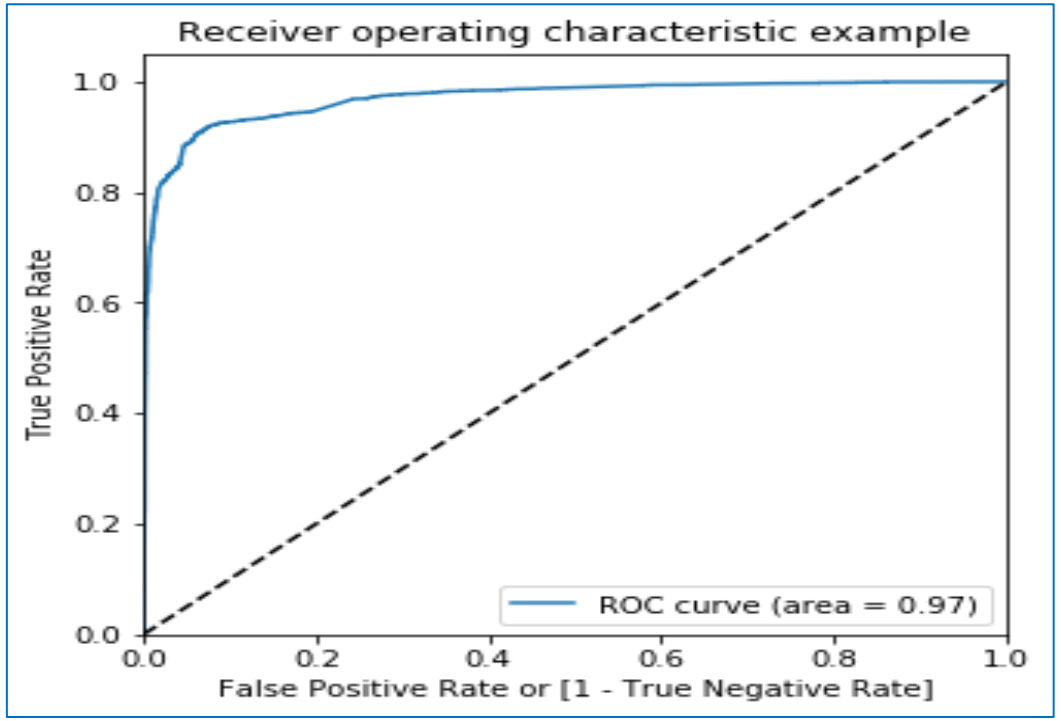
Bivariate Analysis -



- Most number of customers are Management Specialists & conversion rate is also very good. Company may focus more on these leads.
- Customers inquiring about course are majorly unemployed , but the conversion rate is around 50%. Company needs to focus more on converting these leads by giving them assurance about career opportunities after completion of course.
- Working professionals have excellent conversion rate but leads generated are less. Company can attract more working professionals to enroll for the course by letting them know advantages of comleting the course in their professional journey.

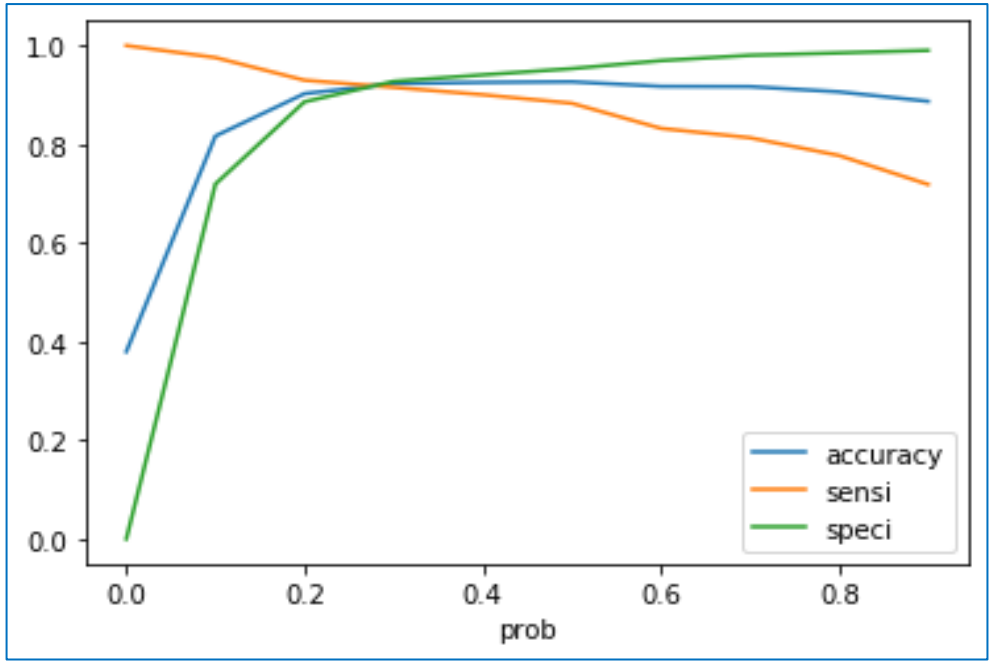
• Plotting ROC Curve

An ROC curve demonstrates several things:
It shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity). The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.



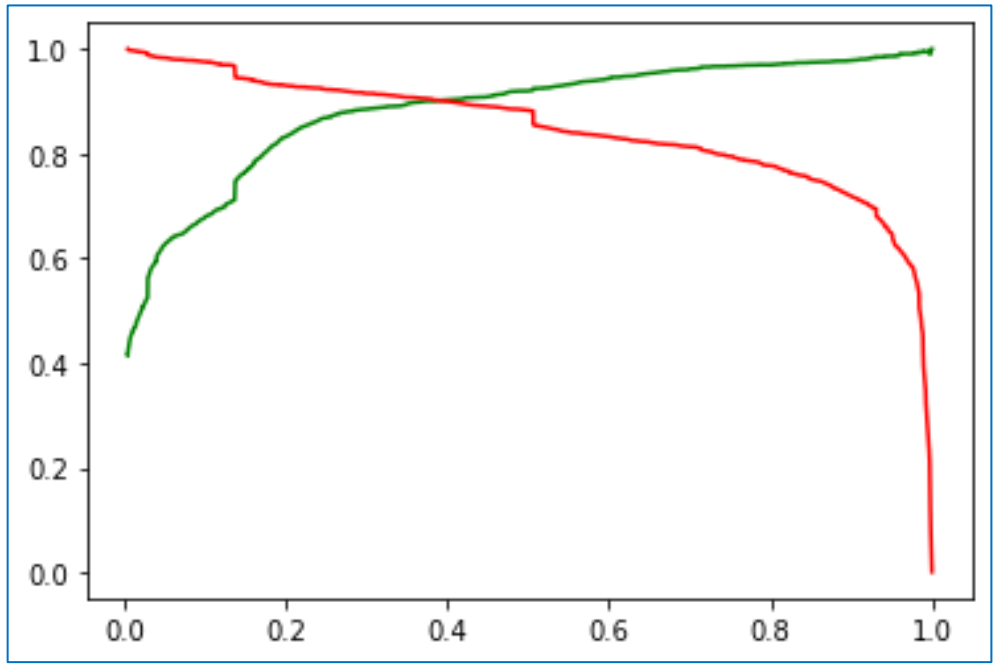
The ROC Curve should be a value close to 1. We are getting a good value of 0.96 indicating a good predictive model.

- **Finding Optimal Cutoff Point**



From the curve above, 0.3 is the optimum point to take it as a cutoff probability.

- **Precision and recall tradeoff**



0.4 is tradeoff between precision & recall .
Thus we can consider any lead with conversion probability of higher than 40 % to be hot lead .

• Final Observations

Let us compare the values obtained for Train & Test:

Train Data:

Accuracy : 92.27% Sensitivity : 91.52% Specificity : 92.73%

Test Data:

Accuracy : 92.78% Sensitivity : 92.13% Specificity : 93.19%

The Model seems to predict the Conversion Rate very well

THANK YOU !