Lead Scoring Case Study

Presented By:

1.Pratiksha Saheb

2.Ami Faldu

3. Safa

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. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. 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%.

- ➢ Now, 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.

Goals of the Case Study

➢ 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.

➢ There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well.

Approach

The steps are following:

- ➢ Reading and Inspecting the data
- ➢ Clean the data(Exploratry Data Analysis)
- ➢ Data Preparation
- ➢ Model Building
- ➢ Model Evaluation

Data Sourcing Cleaning and Preparation

➢ Read the data from the CSV file.

➢ Handling null values and removing column which are having greater than 40% of null value.

➢ Removing Redundant column in the data .

➢ Removing non informative column from the data set .

➢ Imputing Null Values.

➢ Exploratory Data Analysis Approx. conversion rate is 38%.
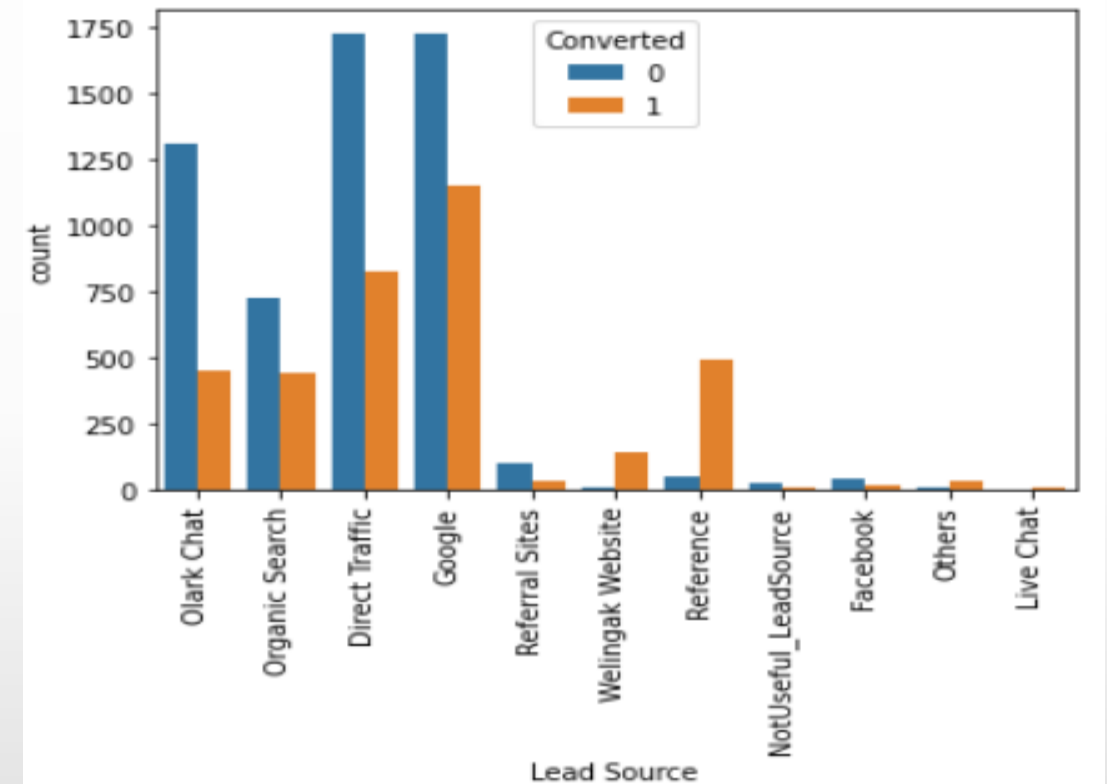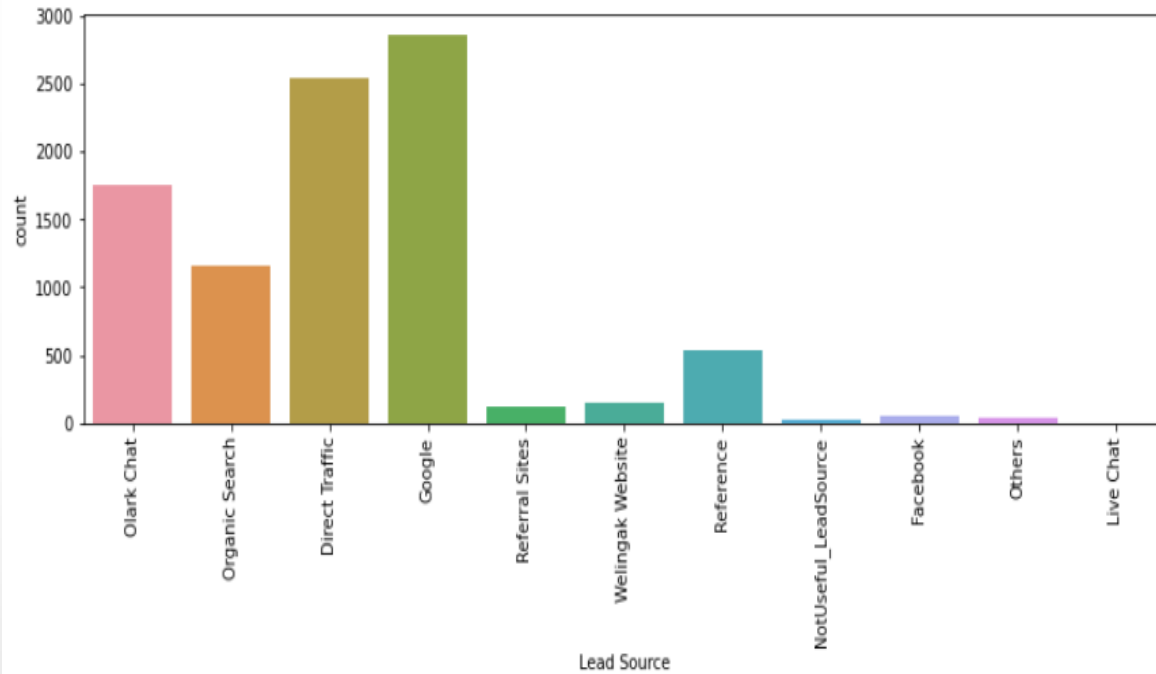
Data Analysis on 'Specilization'
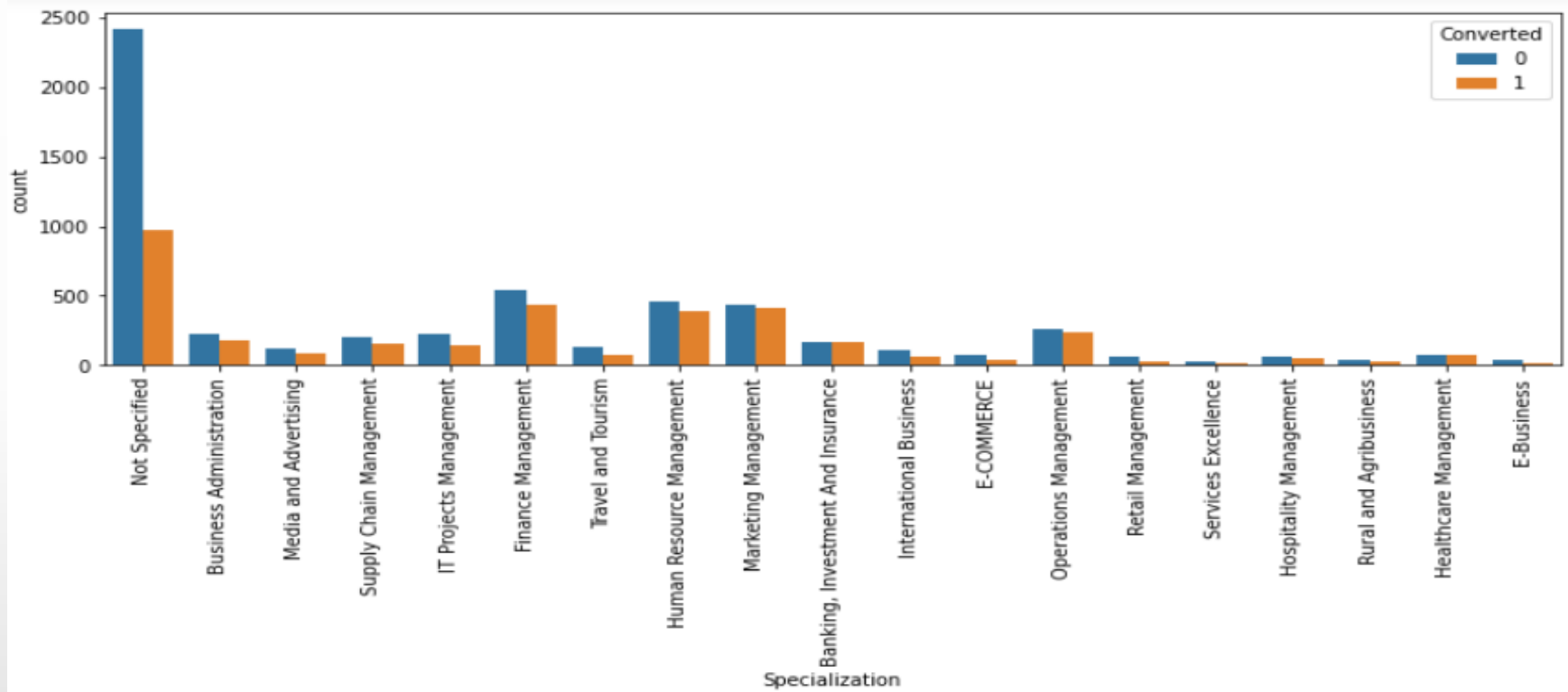
As we can see that specialization with Management in them have higher number of leads as well as leads converted. So this is definitely a significant variable and should not be dropped.
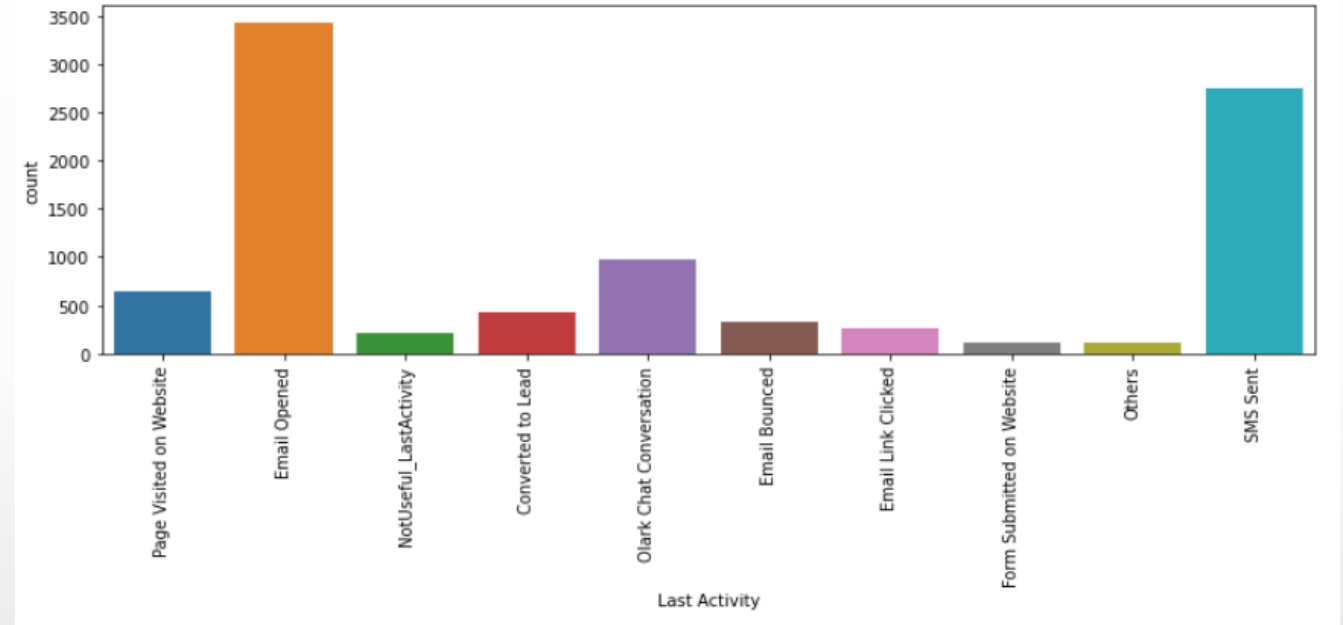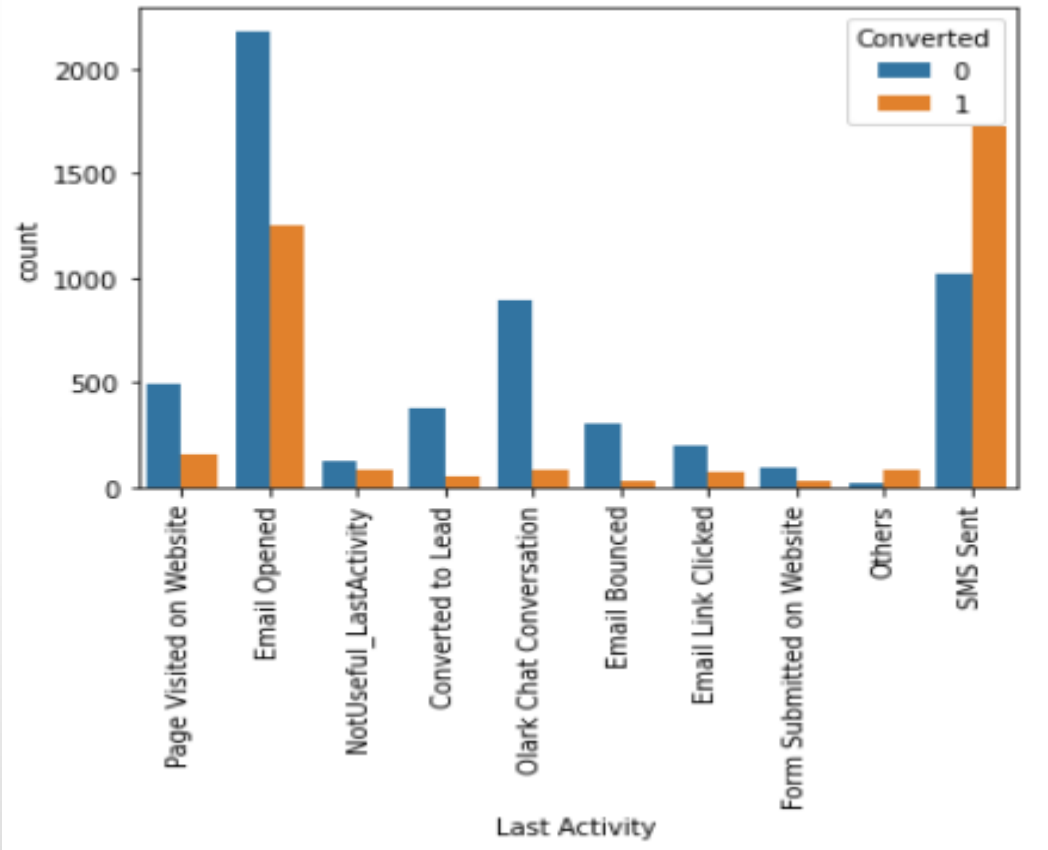
Data Analysis on 'Tags'
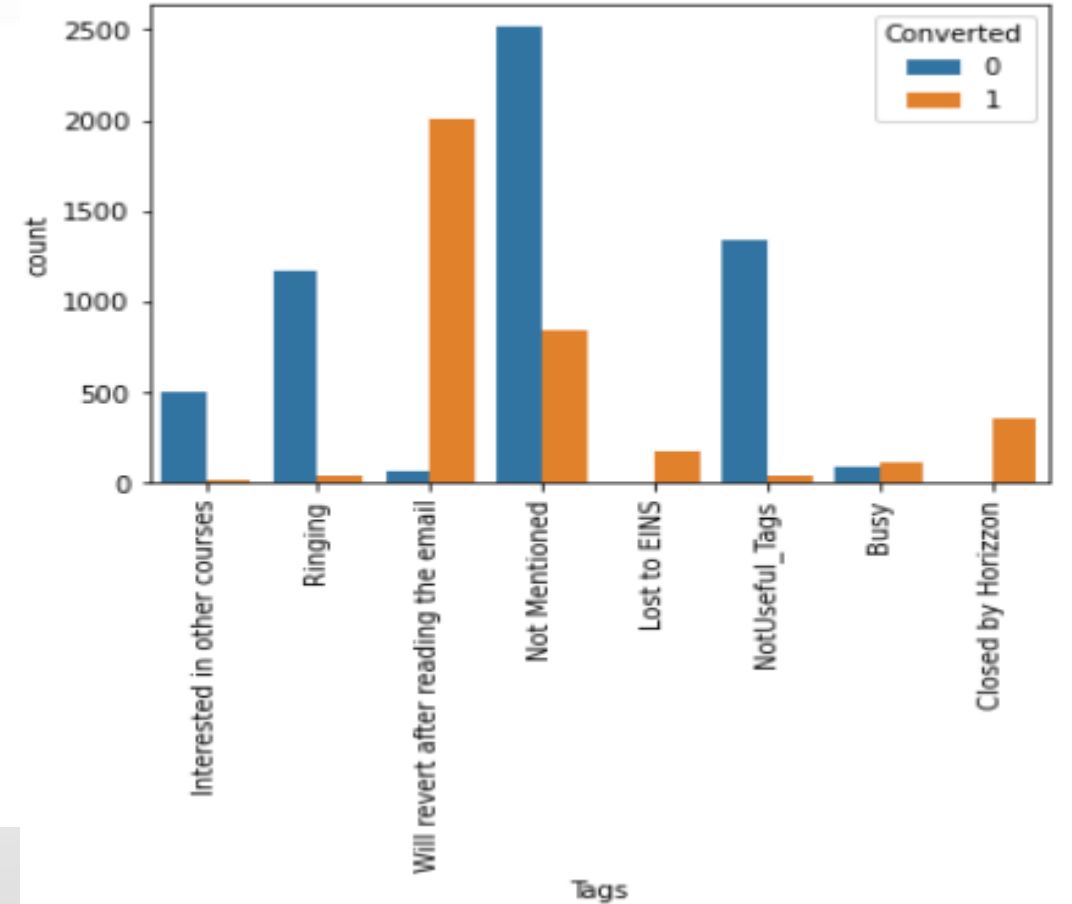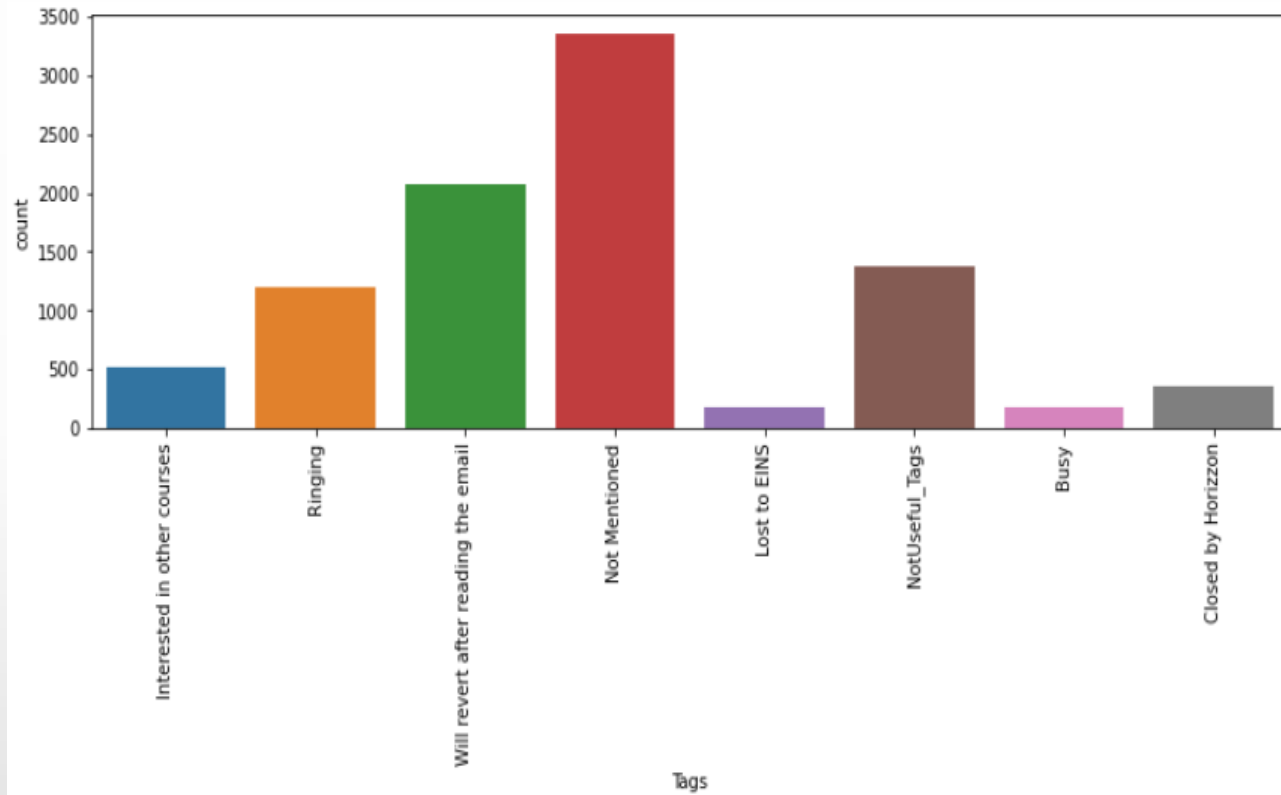
As we can see that 'will revert after reading the email' having most chances for leads conversion
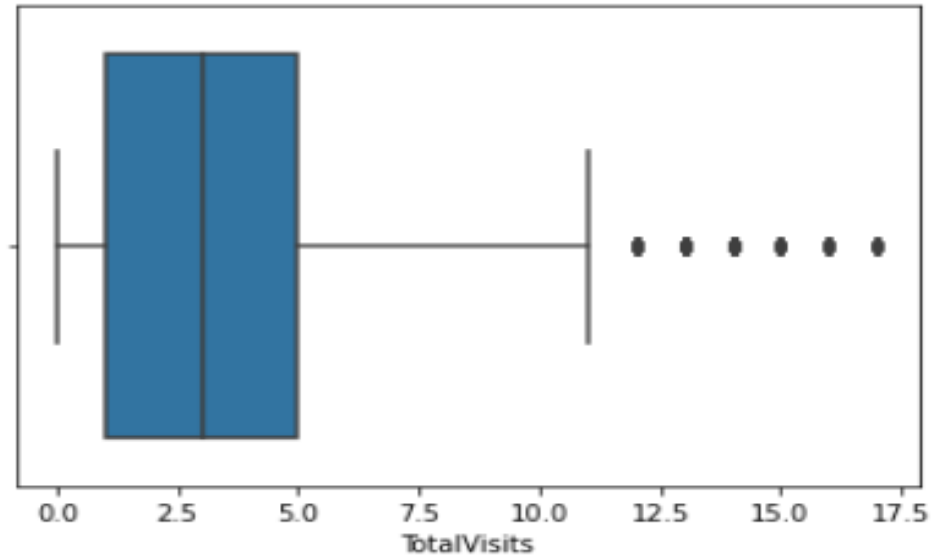
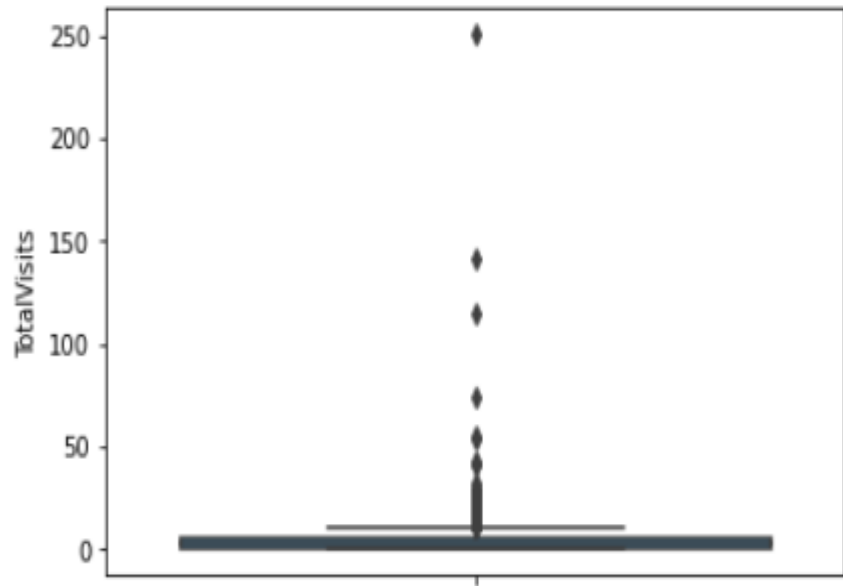# Data Analysis on 'Lead Source'

Maximum number of leads are obtained by Google and Direct Traffic

Olark Chat and Organic Research conversion rate is also high

Data Analysis on 'Last Activity'

as we can see hear SMS sent and Email Opened having high rate of conversion

Outliers Present on the 'Total Visits'

Data Preparation

- Converted Binary Variables in 0 and 1.
- Creating dummy variables for categorical variables.
- Feature Scaling of numeric Data
- Splitting data into train and test sets

Model Building

- Model building using states and RFE
- Determined optimal model logistic Regression machine learning algorithm.
- Calculated Accuracy, Sensitivity, Specificity, and Recall.
- Model Evaluation

Variables Impacting the Conversion Rate

- Total Time Spent on Website

- Lead Origin_Lead Add Form

- Lead Source_Olark Chat

- Lead Source_Others

- Lead Source_Welingak Website

- Last Activity_Email Bounced

- Last Activity_Olark Chat Conversation

- Tags_Closed by Horizon

- Tags_Lost to EINS

- Tags_Not Mentioned

- Tags_Ringing

- Tags_Will revert after reading the email

- Last Notable Activity_Modified

- Last Notable Activity_SMS Sent

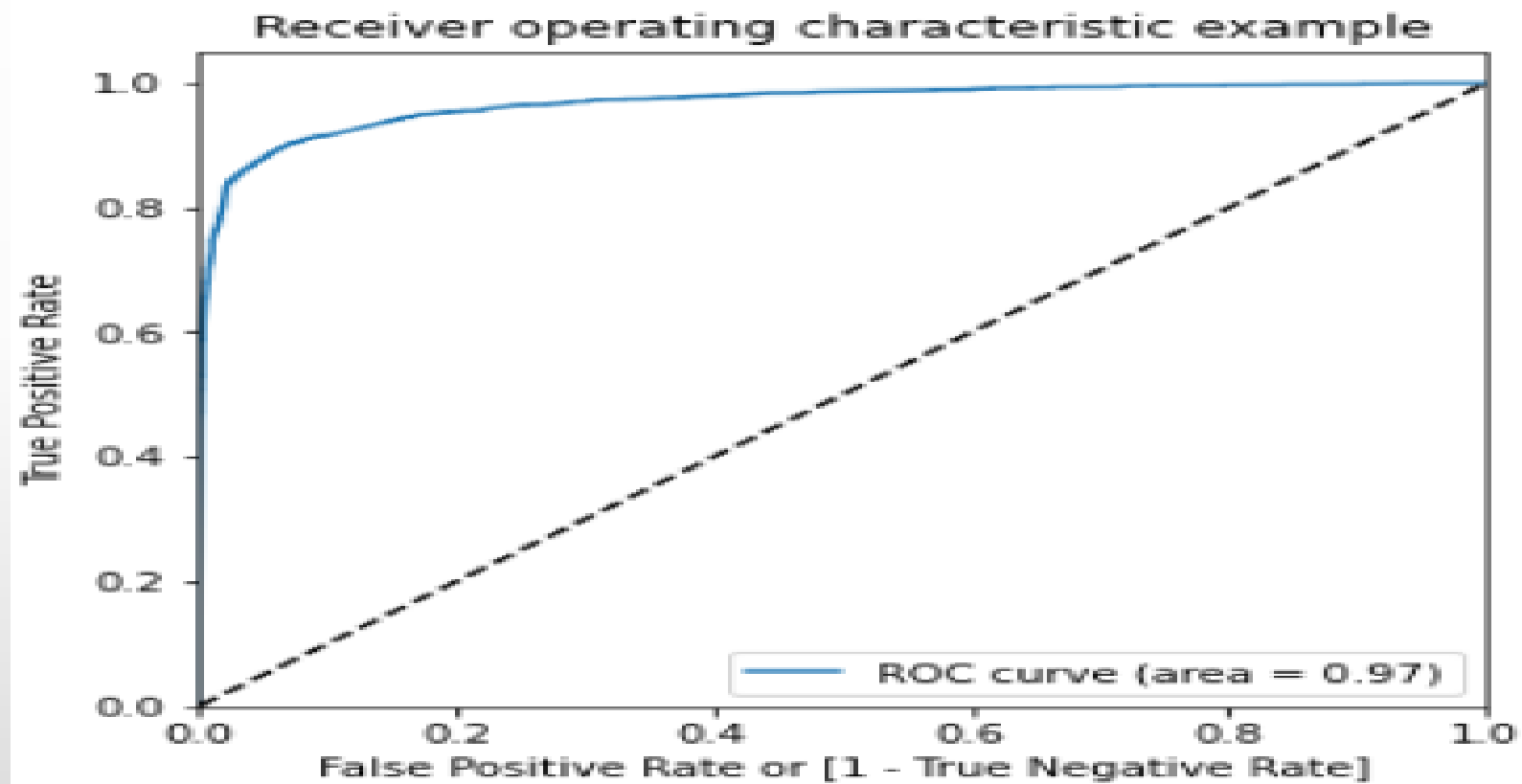# Model Evaluation Sensitivity and Specificity on Train data set
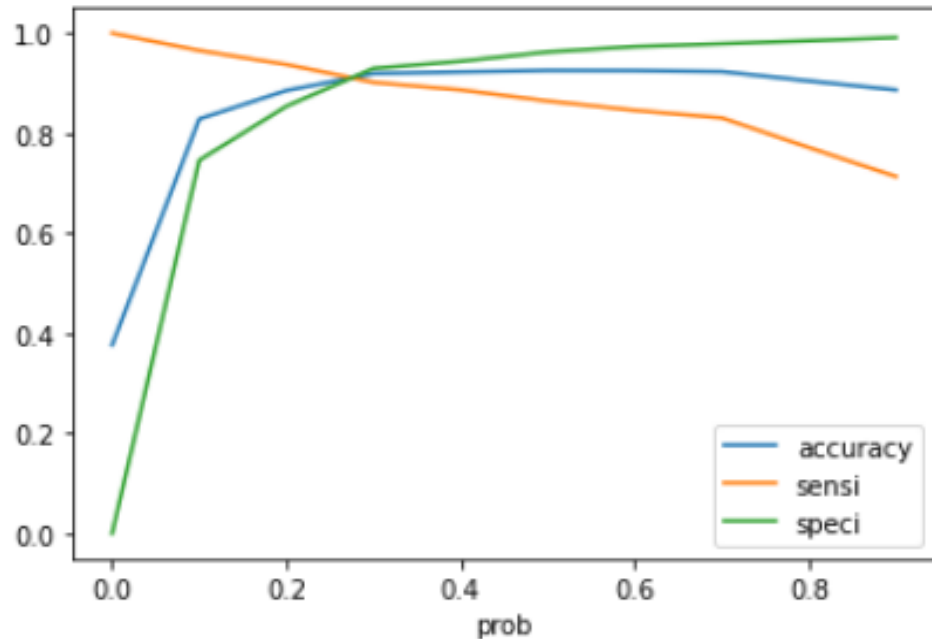
Results based on the Train Set
1. Accuracy = 91.8%
2. Sensistivity(Recall)=90%
3. Specificity=92.9%
4. Precision=88.5%
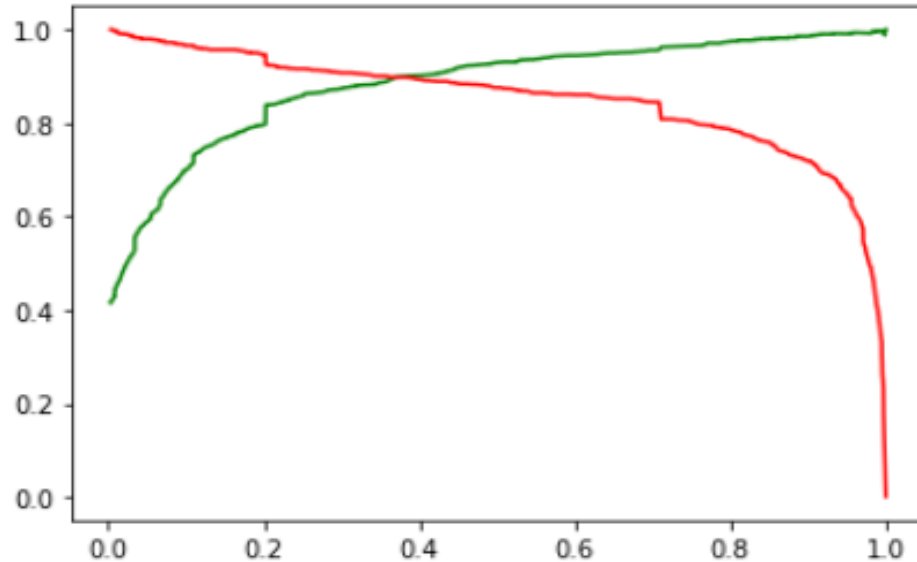5. F1_score=89.3%

Precision –Recall Curve
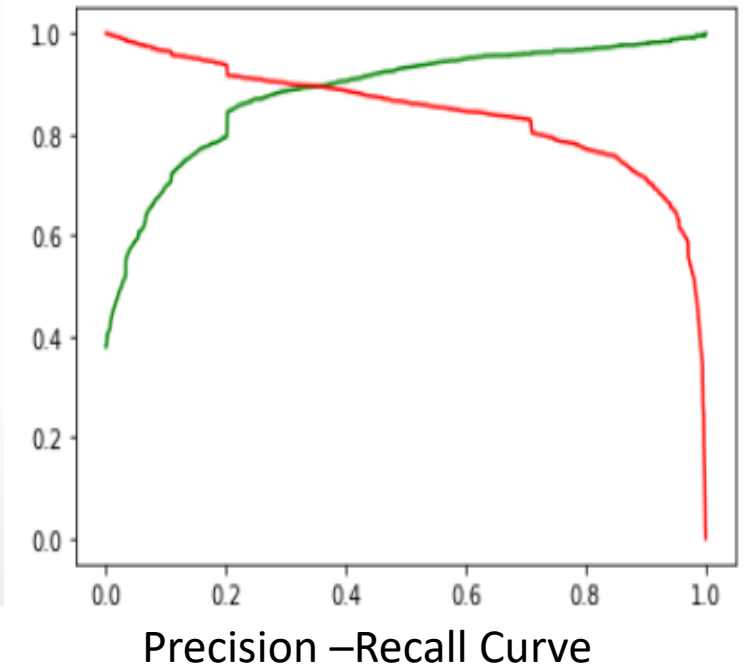
From the curve above, 0.3 is the optimum point to take it as a cutoff probability.
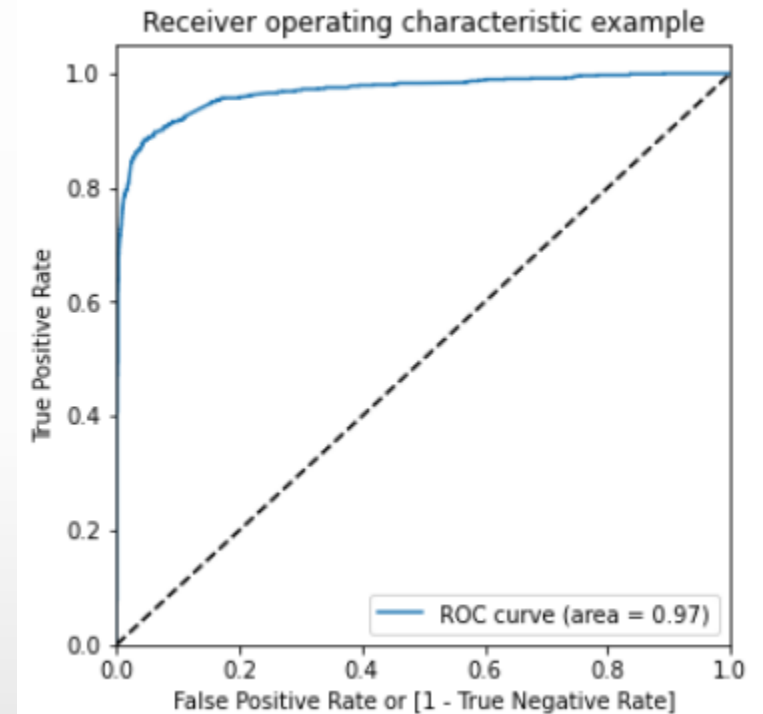
ROC Curve

Model Evaluation Sensitivity and Specificity on Test data set

Precision –Recall Curve

Results based on the Test Set
1. Accuracy = 91.3%
2. Sensistivity(Recall)=90.7%
3. Specificity=91.7%
4. Precision=87.4%
5. F1_score=89%

Results

- Accuracy ,Sensitivity, and Specificity values of test set is 91.3%,90.7%, 91.7% are close to training set.

- While we have checked both sensitivity ,specificity as well as precision and recall matrices of both training and test test sets, and considered the optimal cut off based on sensitivity and specificity for calculating the final prediction .

- Hence overall model seems to be good.

Conclusion

X Education Company needs to focus on following key aspects to improve the overall conversion rate:

➢ Tags_Closed by Horizon, Tags_Lost to EINS and Tags_will revert after reading the email are top three variable which generate high number of leads.

➢ Total time spent on the website by the user creates higher number of leads conversion. Means user engagement on their website helps in higher conversion rate

➢ Increasing sending SMS notifications since this helps in higher leads conversion

➢ Get Total Visits increased by advertising etc. since this helps in higher lead conversion

➢ Increasing the Olark Chat, Welingak Website service, this is helpful for higher leads conversion.

➢ Tag_ringing creating the negative impact on the leads conversion.

➢ Those customers will revert after reading the email showing high leads conversion rate.

Summary

There are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion. First, sort out the best prospects from the leads you have generated. 'Tags _closed by horizon, 'Tags_last to EINS', 'Tags_Will revert after reading the email' which contribute most towards the probability of a lead getting converted. Then, You must keep a list of leads so that you can inform them about new courses, services, job offers and future higher studies. Monitor each lead carefully so that you can tailor the information you send to them. Carefully provide job offerings, information or courses that suits best according to the interest of the leads.

THANK YOU