Lead Scoring Case Study

BY :

➢ **Roshani Pawar**

➢ **Kasula Santosh**

➢ **Navneet**

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. They have process of form filling on their website after which the company that individual as a lead.

▶ 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, this means 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

Business Objective

► To increase the conversion rate, lead X wants to build a lead score model for identification of hot leads
► The target is to achieve 80% conversion rate
► The idea is to build universal model which shall take care of the constraints

Approach

► Importing the data and inspecting the data frame

► Data preparation

► EDA

► Dummy variable creation

► Test-Train split

► Feature scaling

► Correlations

► Model Building (RFE Rsquared VIF and p- values)

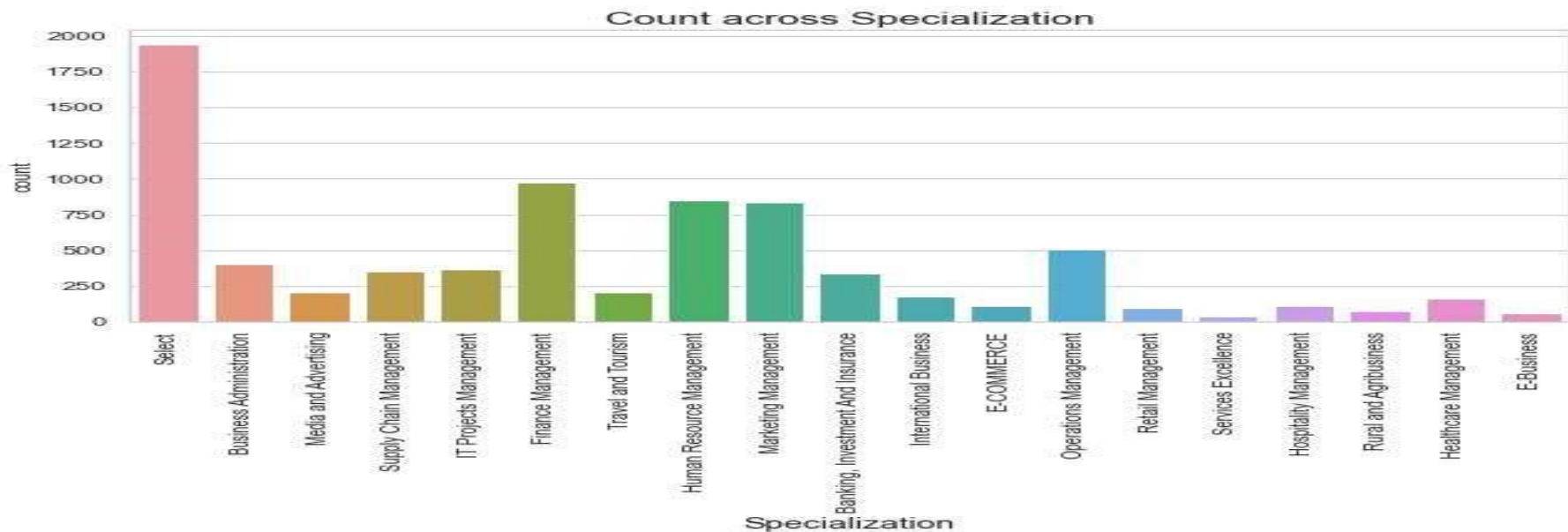► Model Evaluation

► Making predictions on test set

EDA – Data Cleaning

► There are a few columns in which there is a level called 'Select' which is taking care
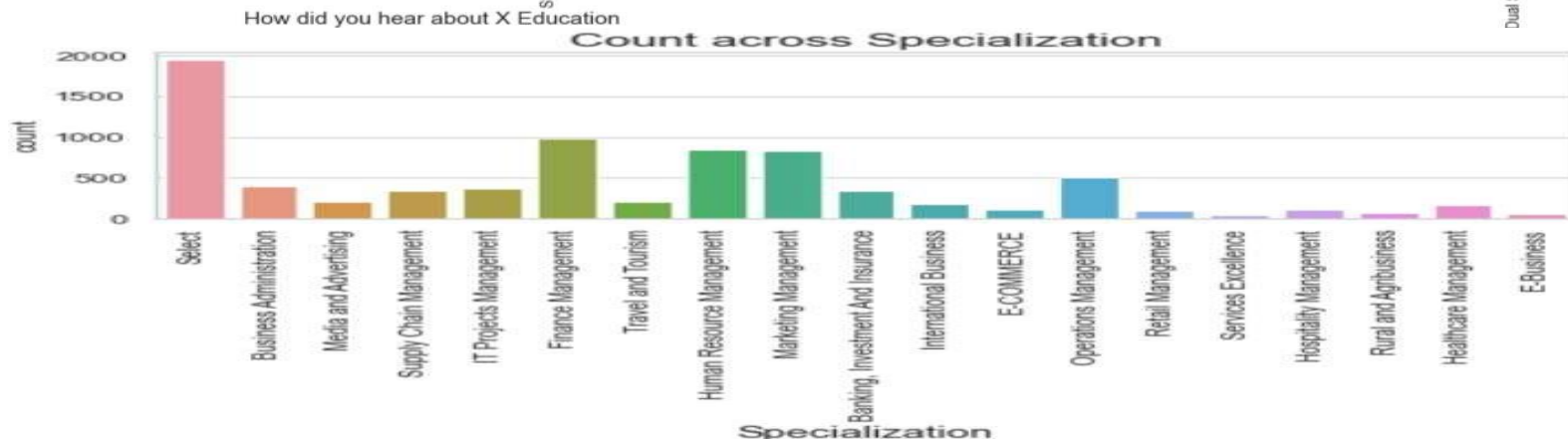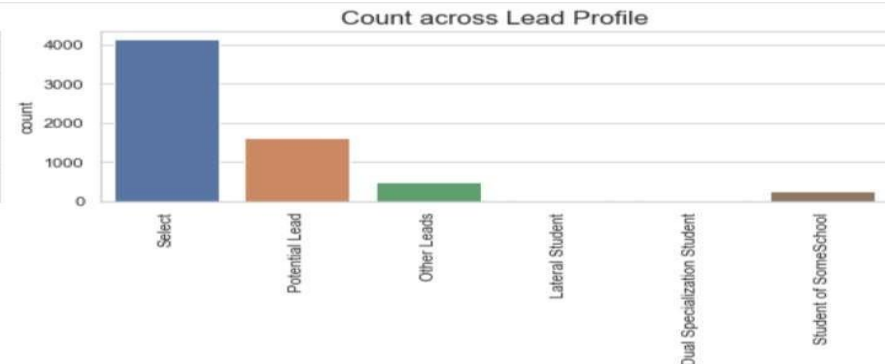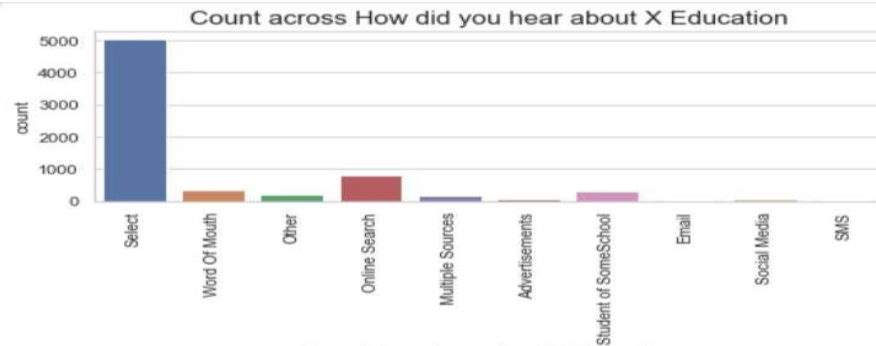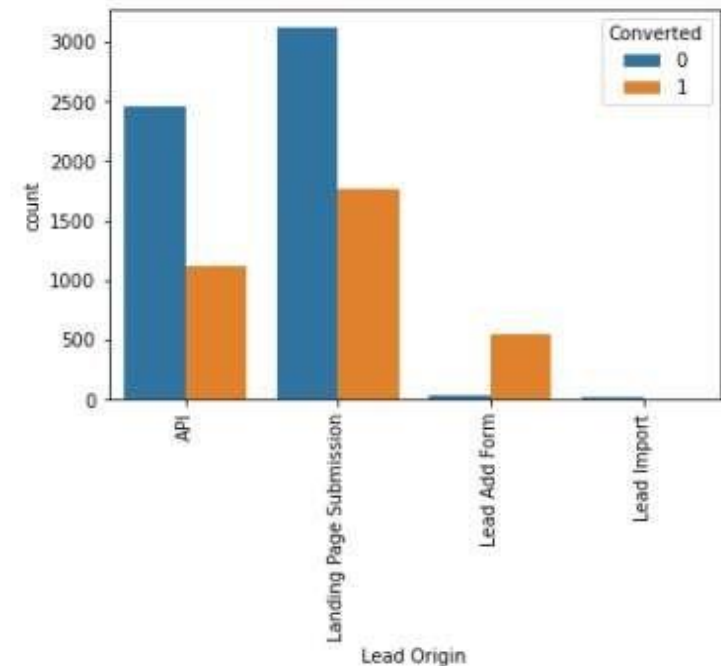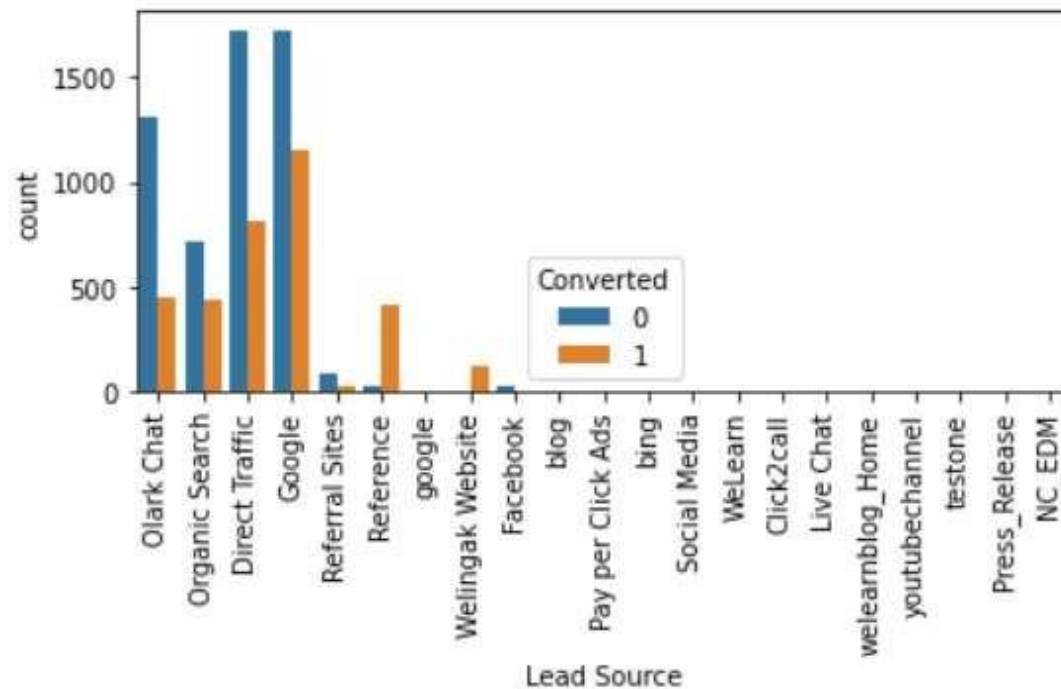
Specialization

Leads from HR, Finance & Marketing management specializations are high probability to convert

Count across Specialization
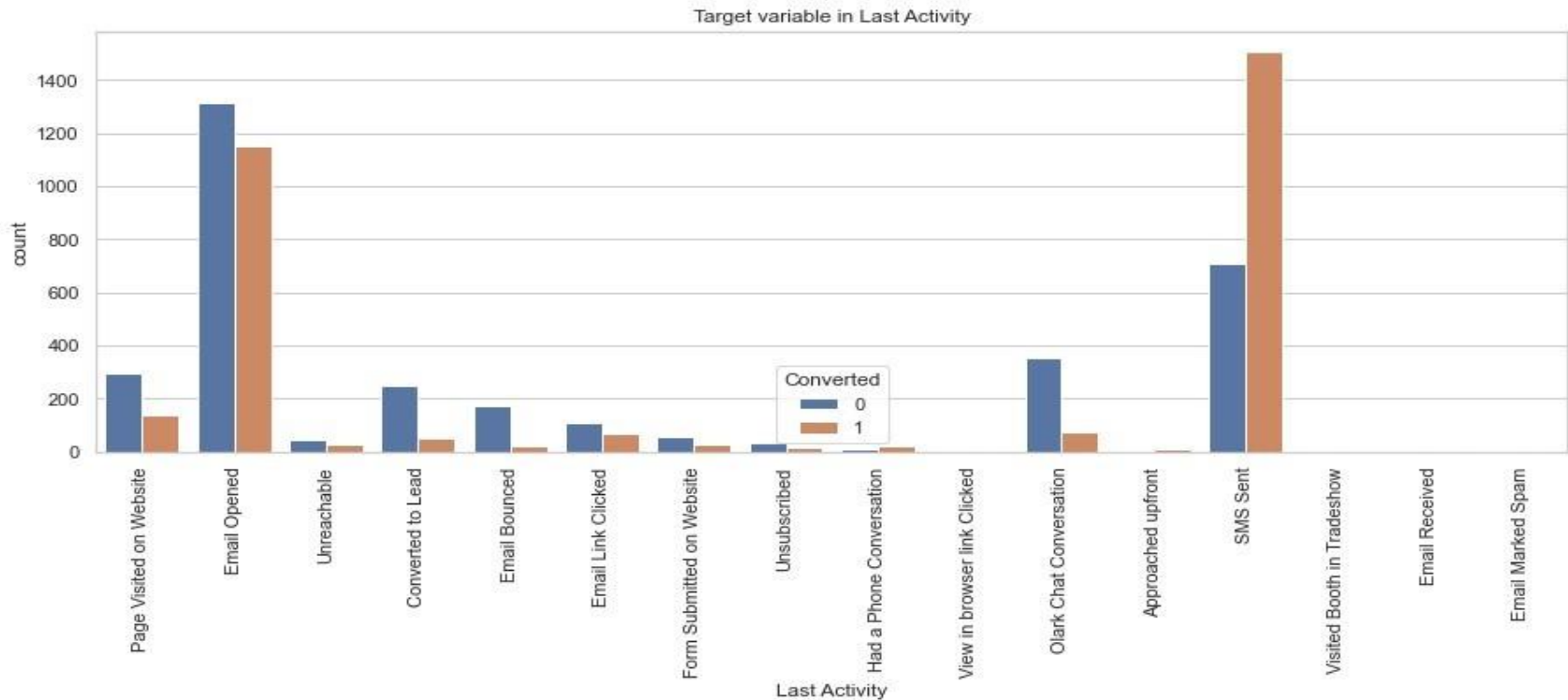
Lead Source & Lead origin

In lead source the leads through google & direct traffic high probability to convert

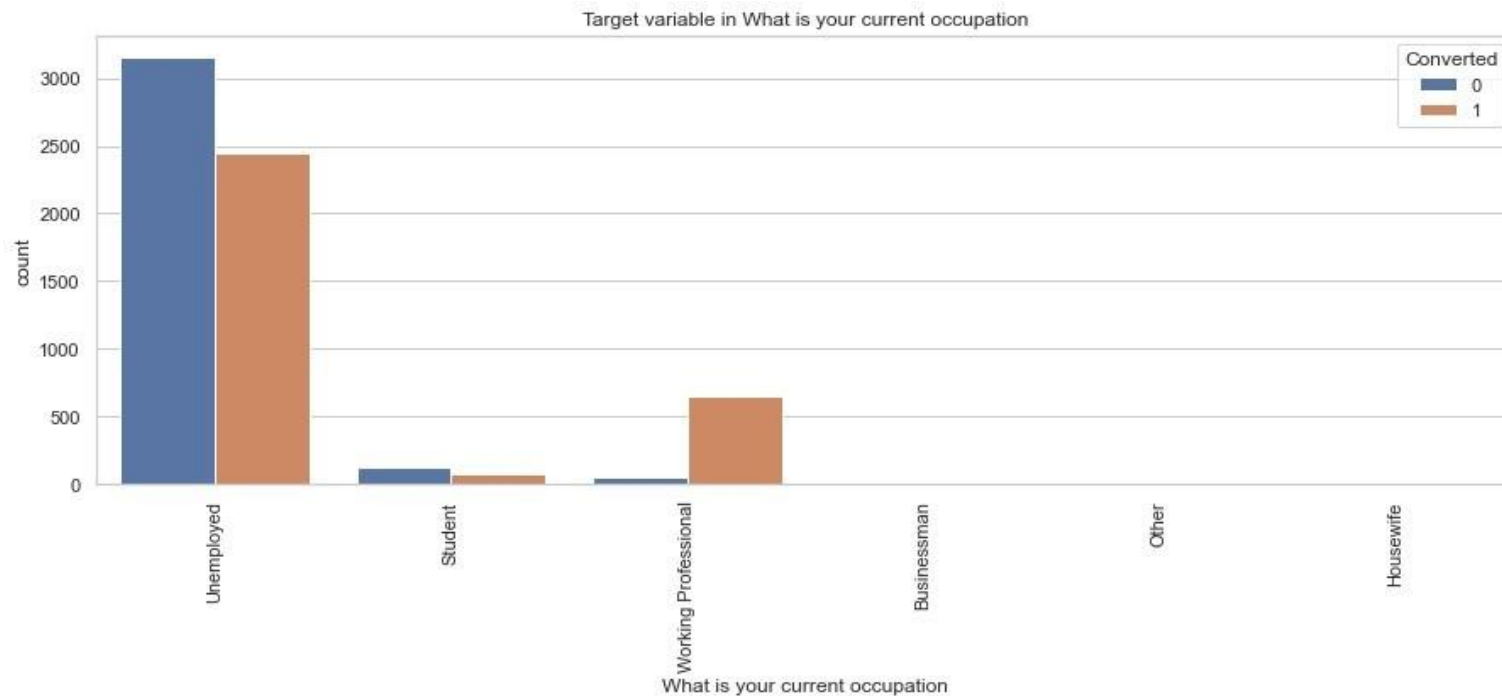Whereas in Lead origin most number of leads are landing on submission

Last lead Activity

Leads which are opening email have high probability to convert, Same as Sending SMS will also benefit.
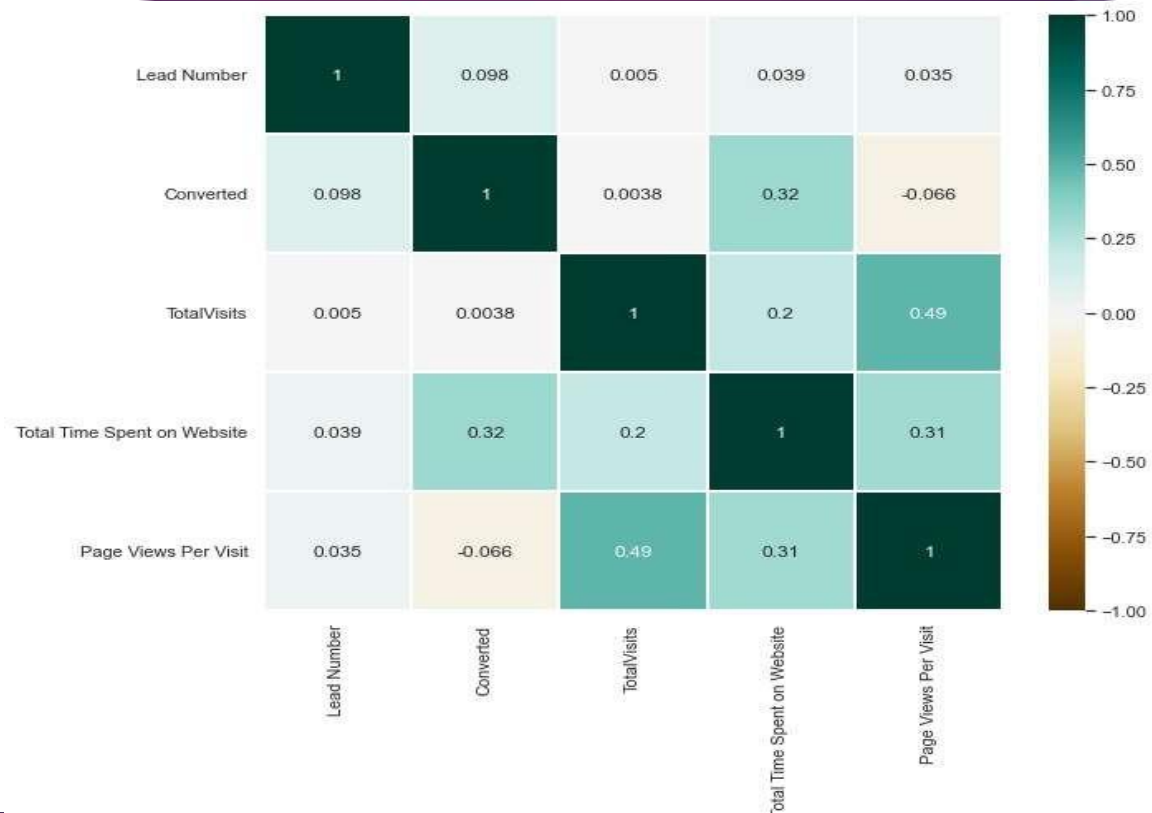

Target variable in Last Activity

Last What is Your Occupation

Leads which are Unemployed are more interested to join the course than others.

Correlation

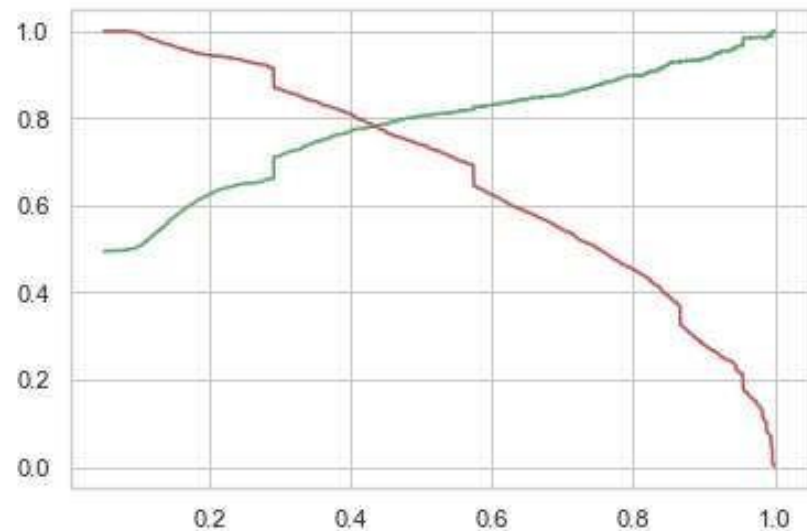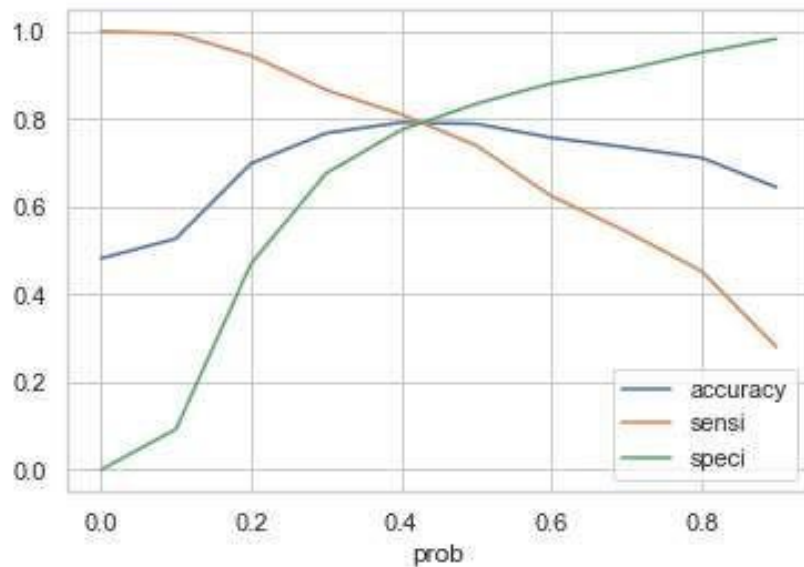There is no correlation between the variables

Model Evaluation

ROC curve

0.42 is the tradeoff between Precision and Recall -
Thus we can safely choose to consider any Prospect Lead with Conversion **Probability higher than 42 % to be a hot Lead**

Observations

Train Data:

Accuracy : 80%

Sensitivity : 77%

Specificity : 80%

Test Data:

Accuracy : 80%

Sensitivity : 77%

Specificity : 80%

Final Features list:

- ► Lead Source_Olark Chat
- ► Specialization_Others
- ► Lead Origin_Lead Add Form
- ► Lead Source_Welingak Website
- ► Total Time Spent on Website
- ► Lead Origin_Landing Page Submission
- ► What is your current occupation_Working Professionals
- ► Do Not Email

Conclusion

► The lead from the API and the landing page is having high conversion rate (30 – 35%) when compared with the lead import

► Maximum conversion is from welingak and reference

► Google and direct traffic has the maximum generation of leads

► Time spent on website is directly proportional to the conversion rate

► Most common last activity is email opened. highest rate = SMS Sent. Max are unemployed. Max conversion with working professional.

Thanks 🙂