Lead Scoring Case Study Using Logistic Regression

SUBMITTED BY :

RAKESH

PRITHA

POORNIMA

Contents

Problem Statement

▶ An education company , X education sells online courses to industry professionals. On any given day , many professionals who are interested in the courses land on their websites and browse for their courses. They have process including form filling on their website after which the company will consider the candidate as a lead.

▶ Once these leads are acquired , employees from the sells team starts making calls, writing mails etc. Through these processes, some of the leads get converted while most do not.

▶ The typical leads conversion rate in X education company is 30%. This means if they acquire 100 leads a day , only 30 of them are converted. To make this process more efficient, company wants to identify more potential leads named Hot Leads.

▶ If they successfully identify these leads, the conversion rate should goes up as the sales team will now be focusing on the potential leads rather than making calls to everyone.

Business Objectives

▶ The company wants to build a model to give a score to every lead between 0-100. So that they can easily identify the hot leads and increase their conversion rate as well.

▶ The CEO wants the conversion rate of 80%

▶ They want the model to be able to handle future constraints as well as Peak time action required , how to utilize full man power and after achieving target what should be the approaches.

Problem Approach

- Importing the data and Inspecting the dataset.

- Data preparation

- EDA

- Dummy Variable Creation

- Train-Test split

- Feature Scaling

- Correlations

- Model Building by VIF and p-values

- Model Evaluation
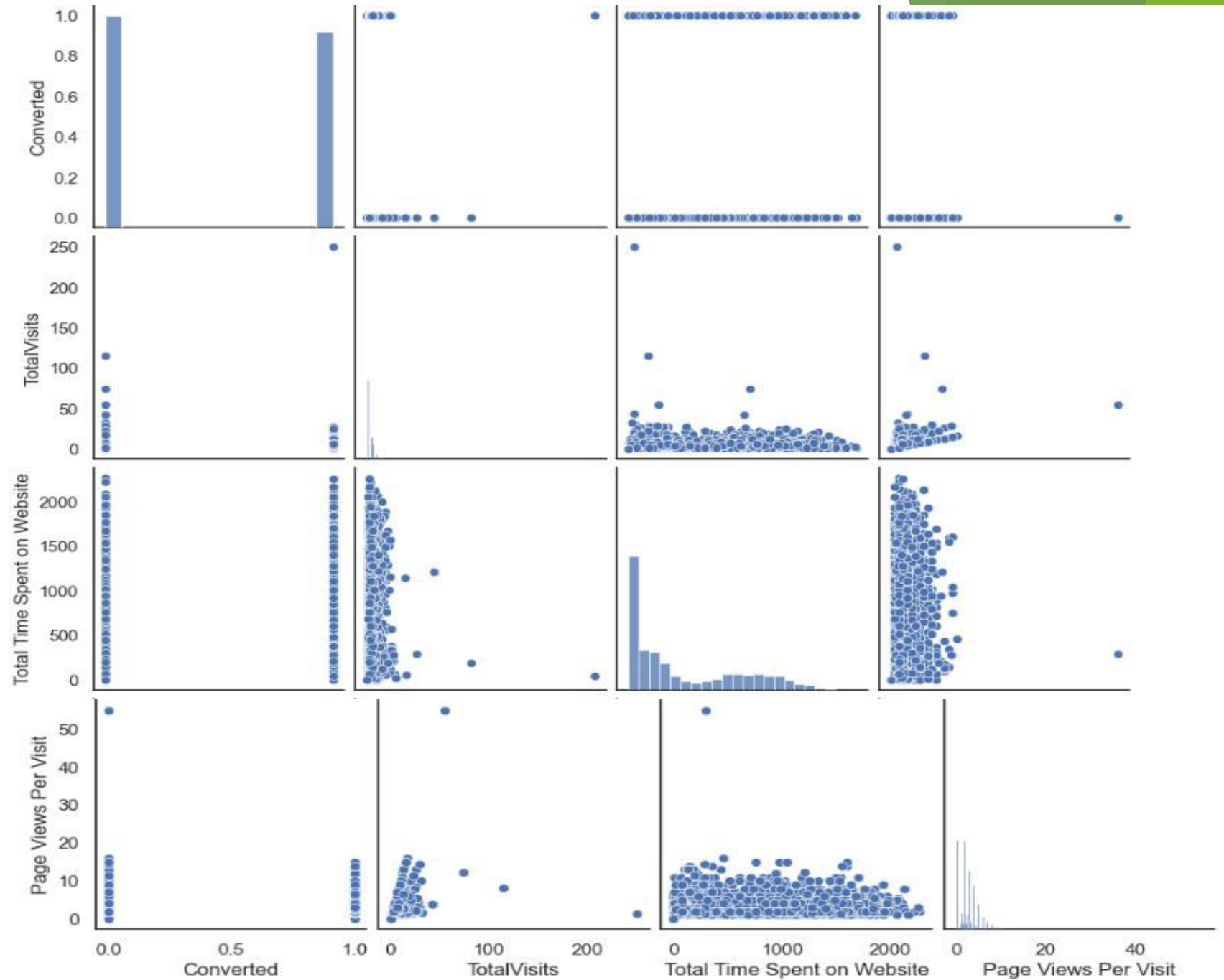
- Making Prediction on Test set

EDA- Data Cleaning

▶ There are three columns where the level is 'Select', which is of no use.
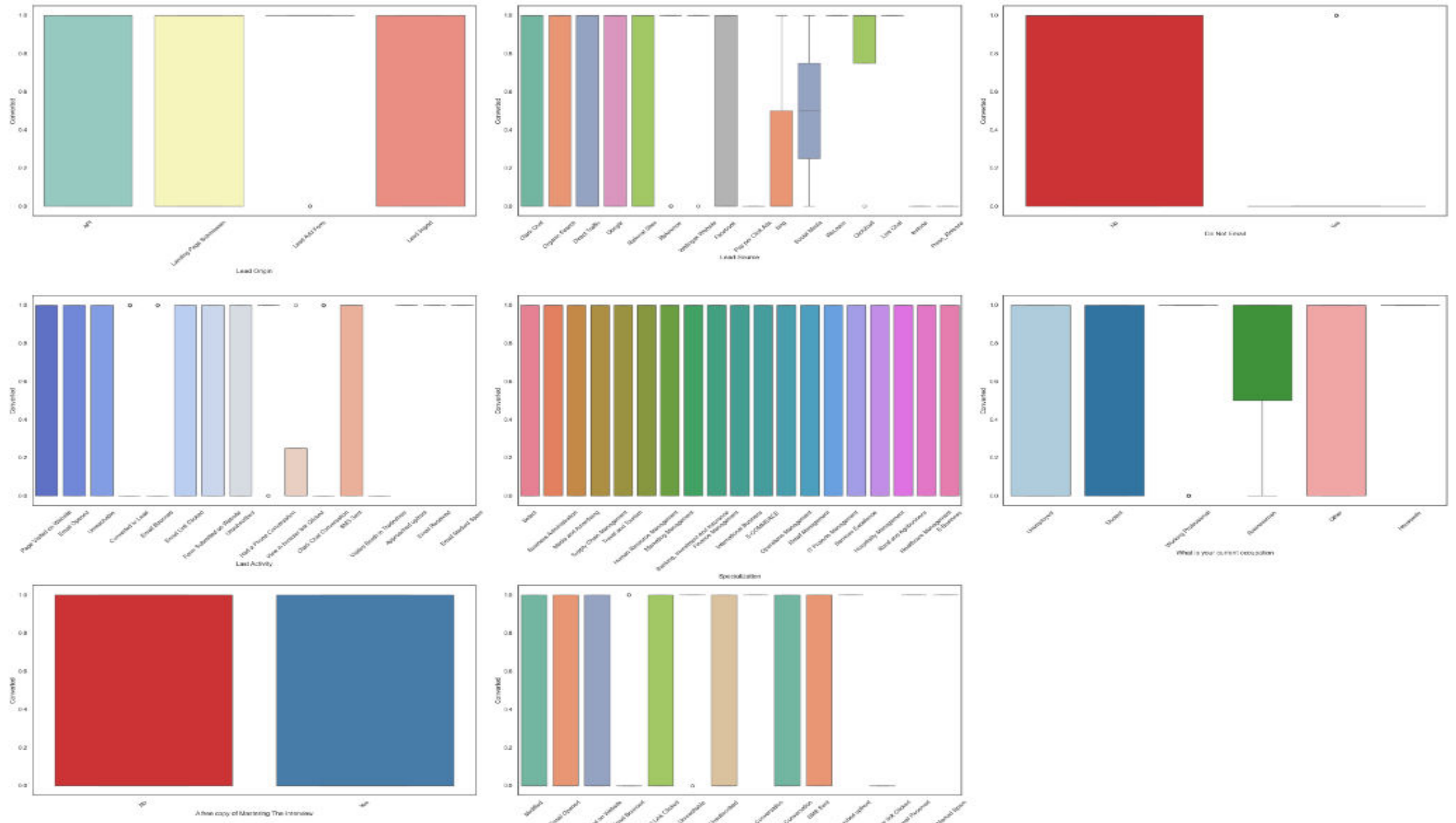
▶ So we need to drop them carefully.
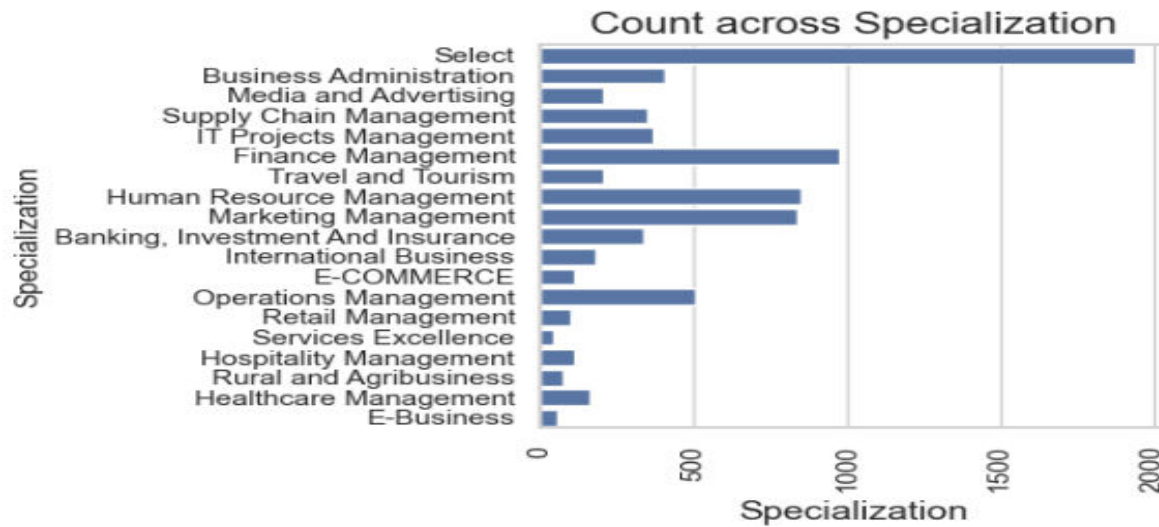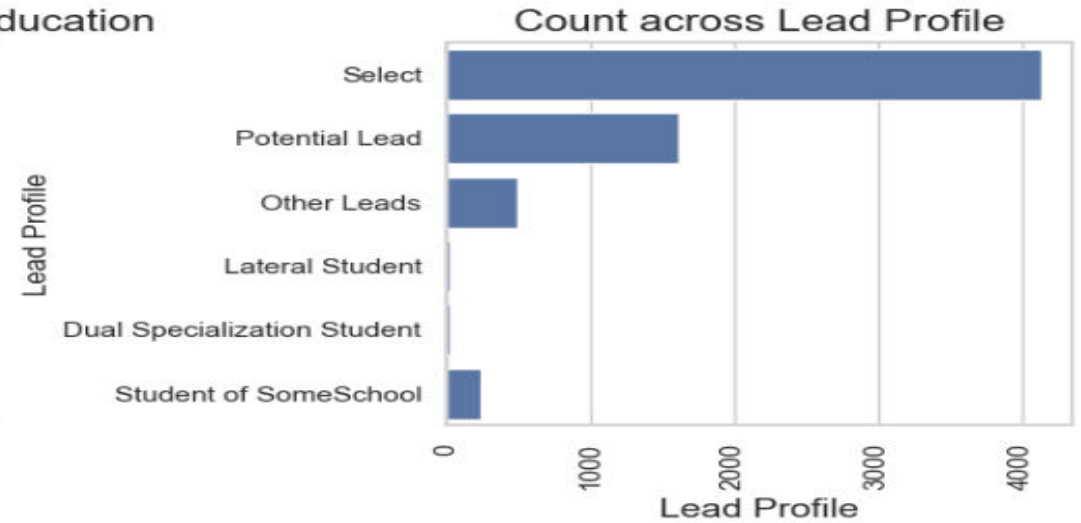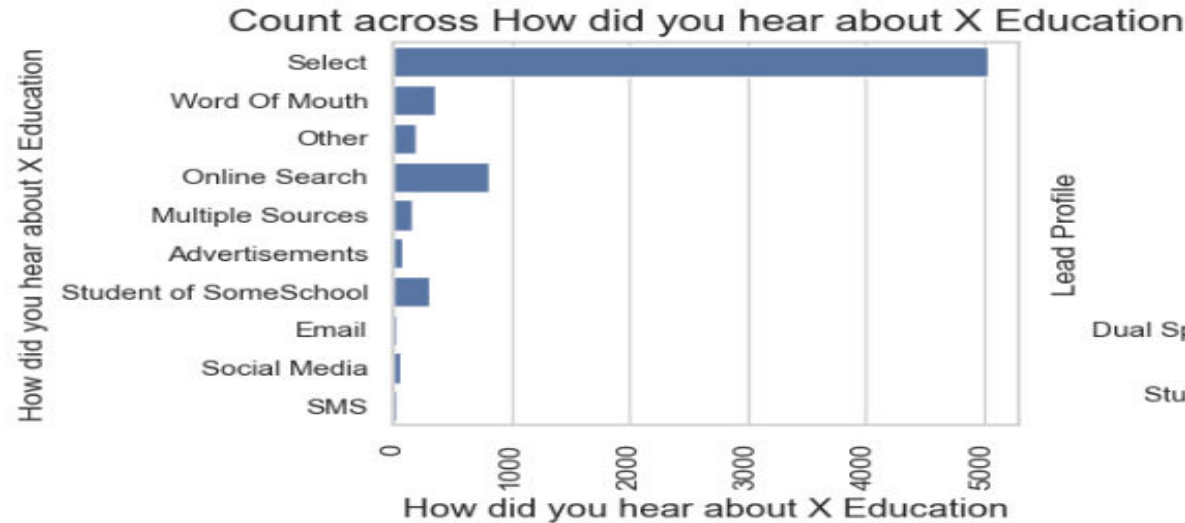
EDA-Numerical Variables
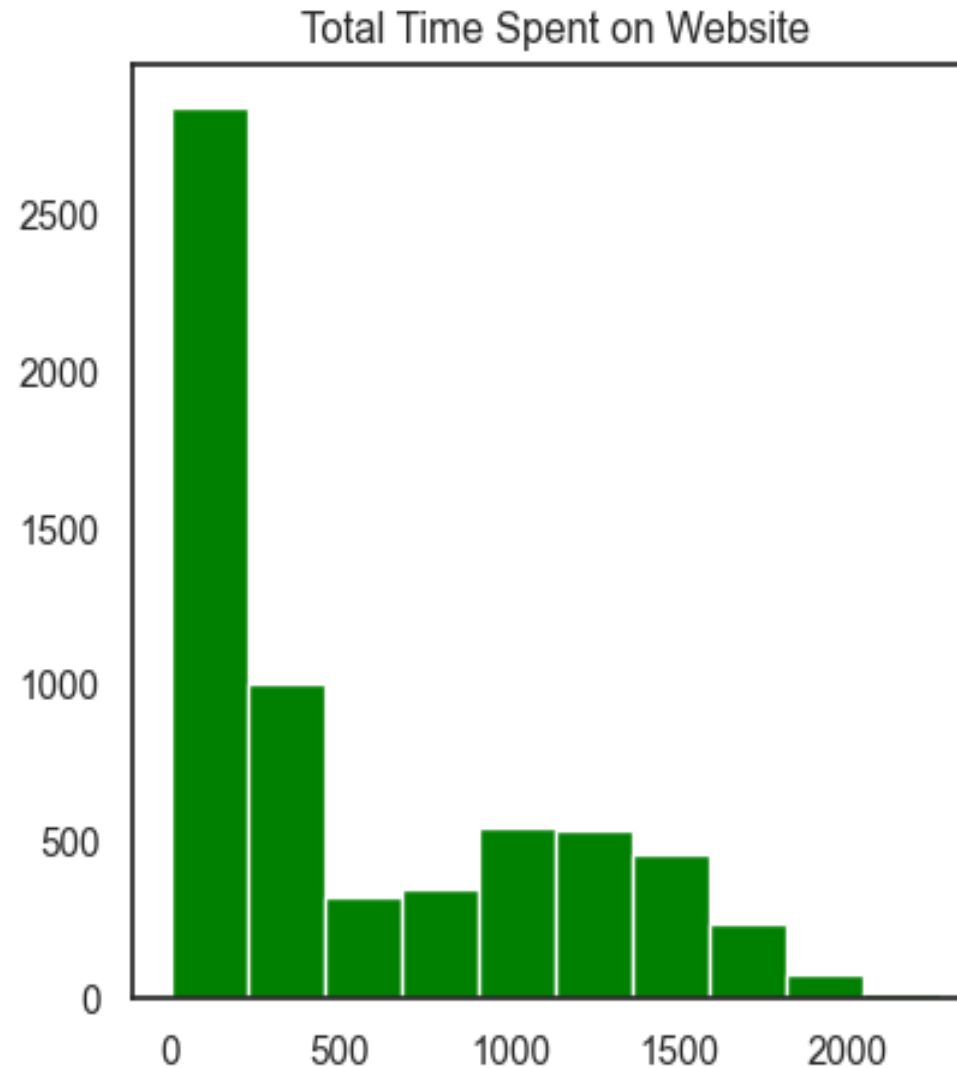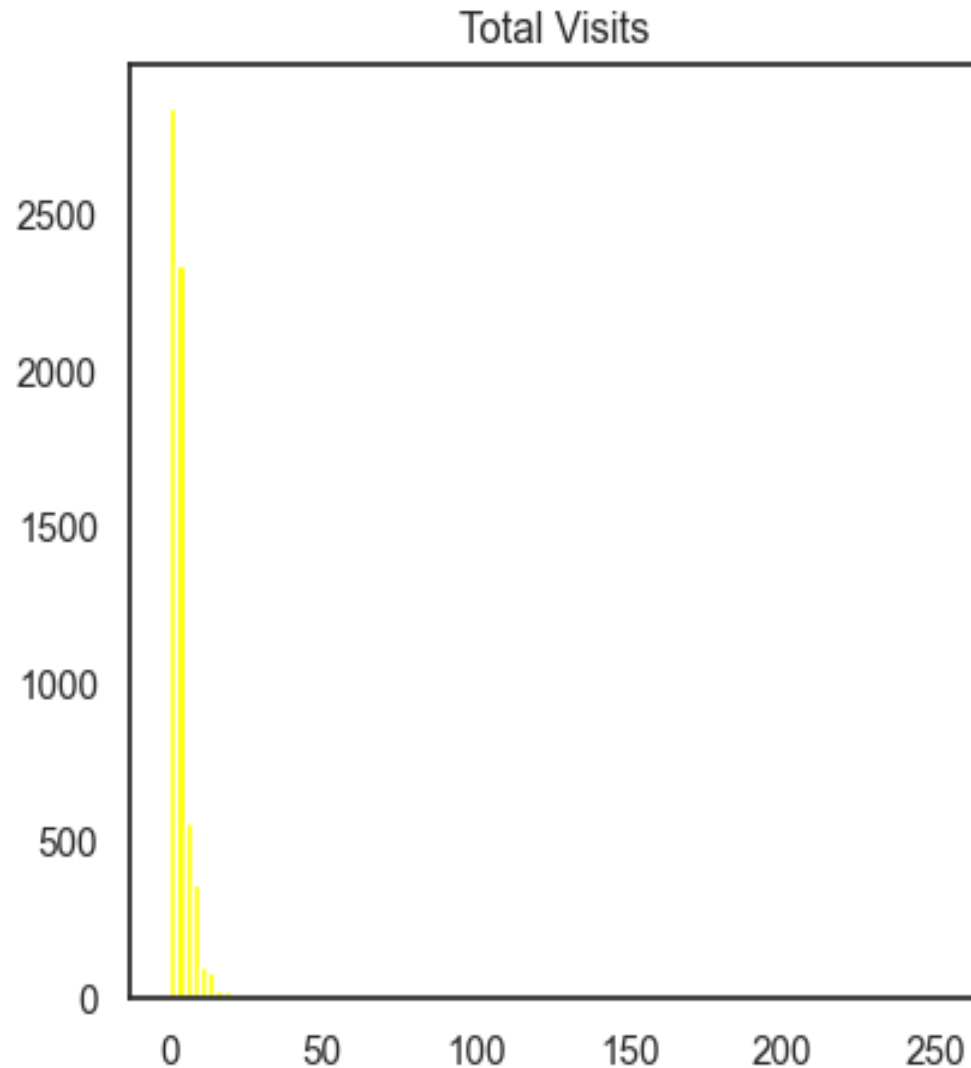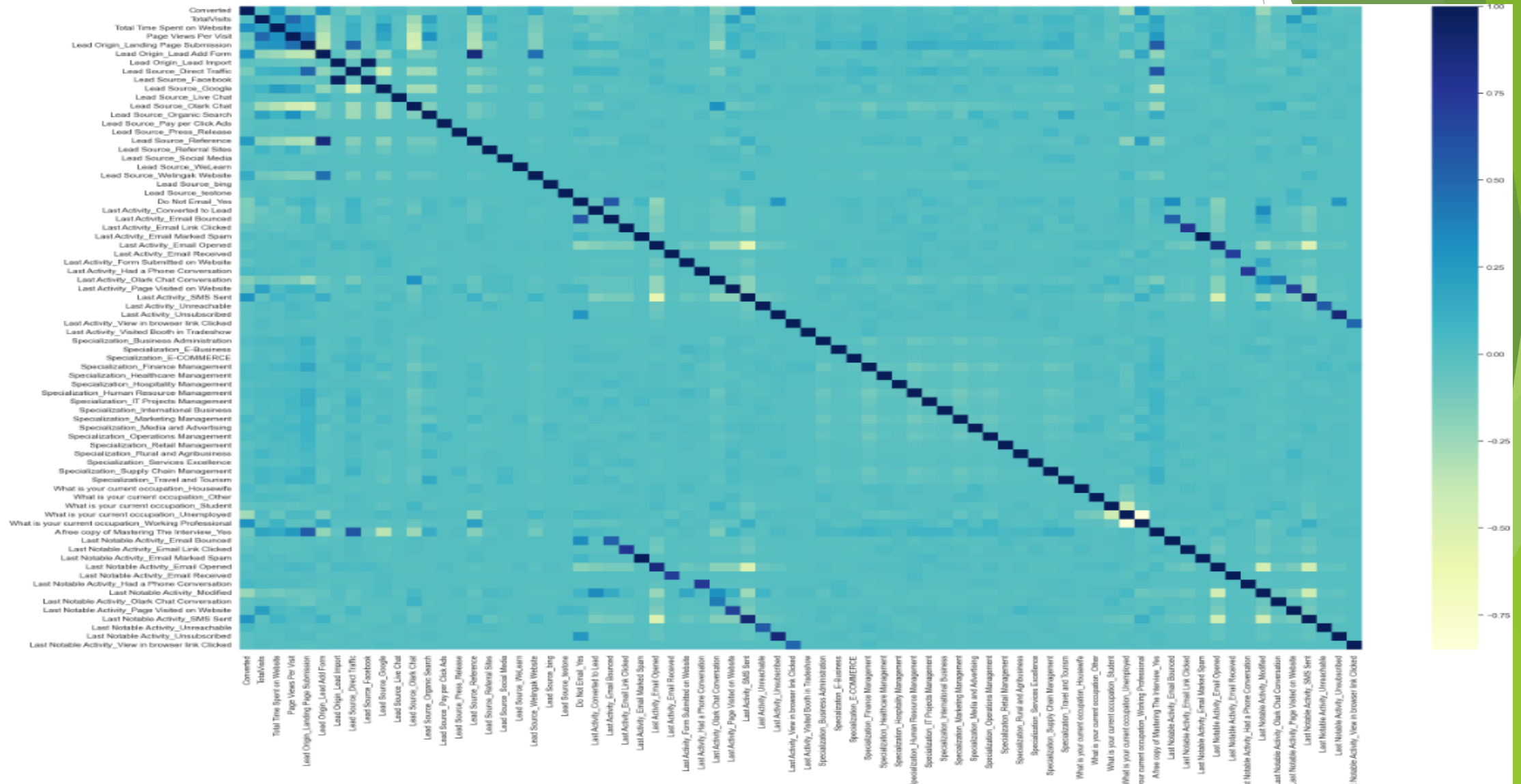
EDA-
Correlation
between all
Numerical
Variable

# EDA-Visualizing Categorical Column

Correlation- Since the number of variables are pretty high here , but from the above correlation diagram we can say that there are many variables that are highly correlated to each other.
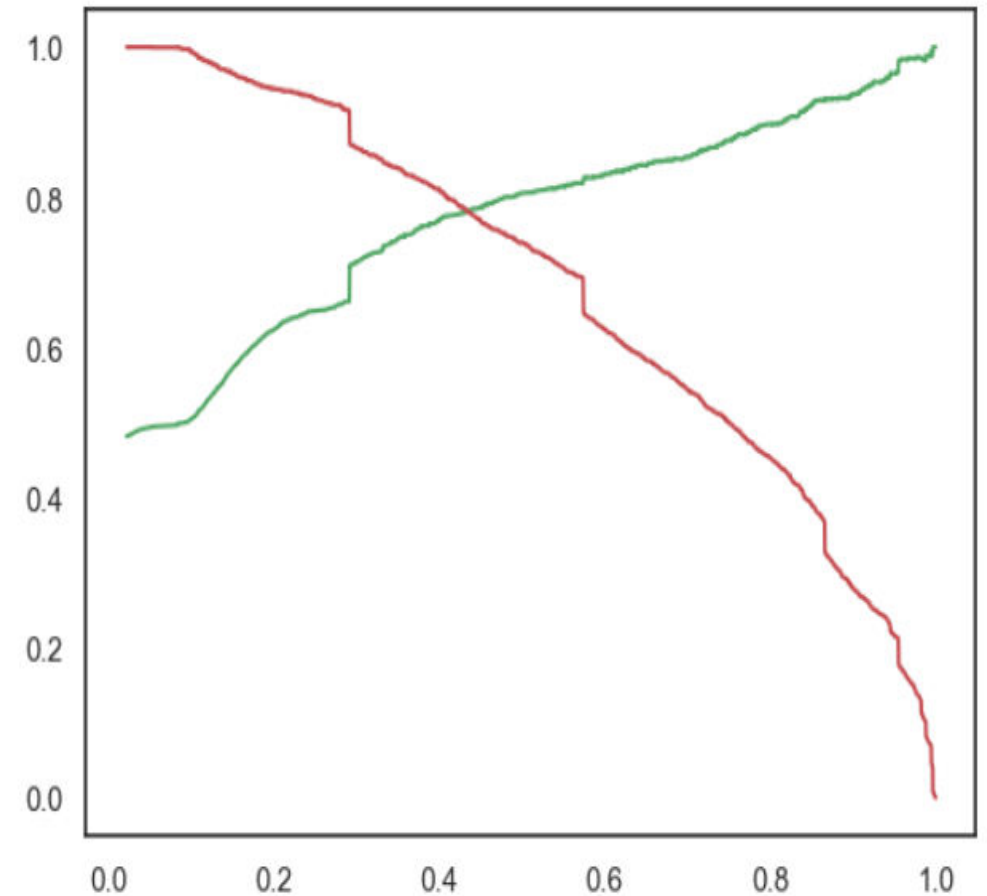
Model Evaluation – 0.42 is the trade off between Precision and Recall. Thus we can obviously tell that any lead with conversion probability higher than 42% must be a Hot Lead.
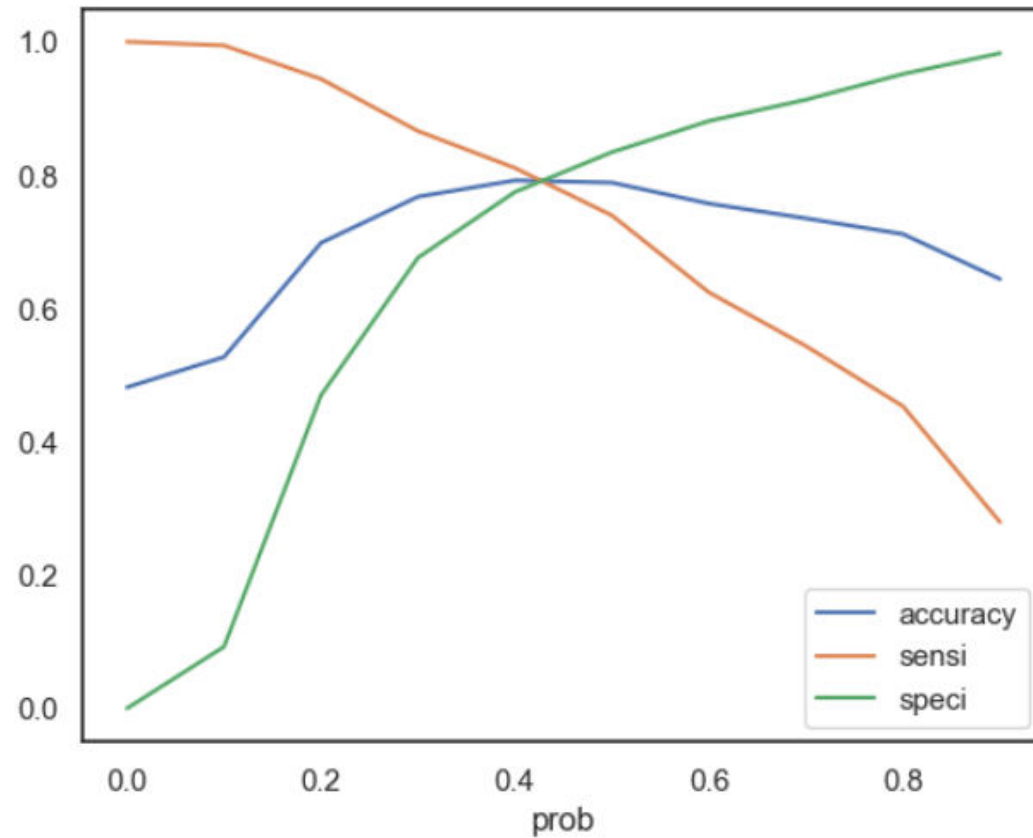
As we can see that around 0.42 or 0.43, So let's choose 0.42 as our cutoff now.

Observations

Train Data :

Accuracy: 79%

Sensitivity: 79.3%

Specificity:78.8%

Test Data:

Accuracy:78.4%

Sensitivity:77.9%

Specificity:78.9%

► ## Final Features List:

- What is your current occupation_Unemployed
- Total Time Spent on Website
- Total_Visits
- Last Activity_SMS Sent
- Lead Origin_Lead Add Form
- Lead Source_Olark Chat
- Lead Source_Welingak Website
- Do Not Email_Yes
- What is your current occupation_Student
- Last Activity_Had a Phone Conversation
- Last Notable Activity_Unreachable

Conclusion

▶ 1.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, we nurtured the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.

▶ 2. First, sort out the best prospects from the leads you have generated. 'TotalVisits' , 'Total Time Spent on Website' , 'Page Views Per Visit' which contribute most towards the probability of a lead getting converted.

▶ 3. Then, we must keep a list of leads handy so that we can inform them about new courses, services, job offers and future higher studies. Monitor each lead carefully so that we can tailor the information carefully , provide job offerings, information or courses that suits best according to the interest of the leads. A proper plan to chart the needs of each lead will go a long way to capture the leads as prospects.

▶ 4.Some factors on which we focus carefully are from Lead Source (Lead Source_Olark Chat, Lead Source_Welingak Website), What is your current occupation(Student, Unemployed),Total Time Spent on Website , TotalVisits, Last Activity_SMS Sent, Lead Origin_Lead Add Forum, Do Not Email_Yes, Last Activity_Had a Phone Conversation, Last Notable Activity_Unreachable.

▶

Thank You