Lead Score Assignment

- Submitted by Group:
- Manika Gupta
- Satvik Praveen
- Animesh Gupta

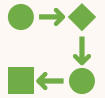
Problem Statement

An education company named X Education sells online courses to industry professionals.

Many professionals who are interested in the courses land on their website and browse for courses.

Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos.

Once these leads are acquired, employees from the sales team start making calls, writing emails, etc.

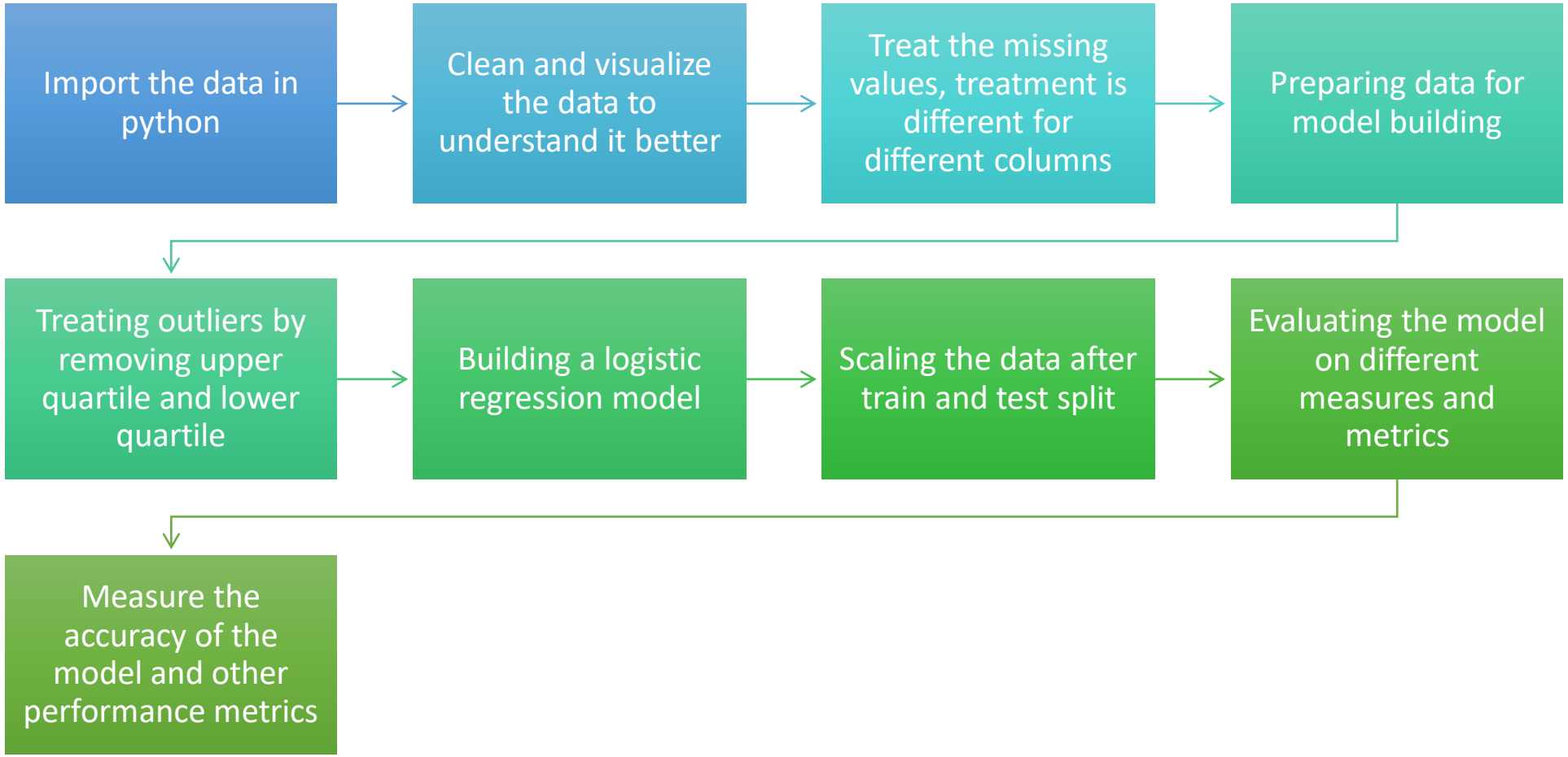


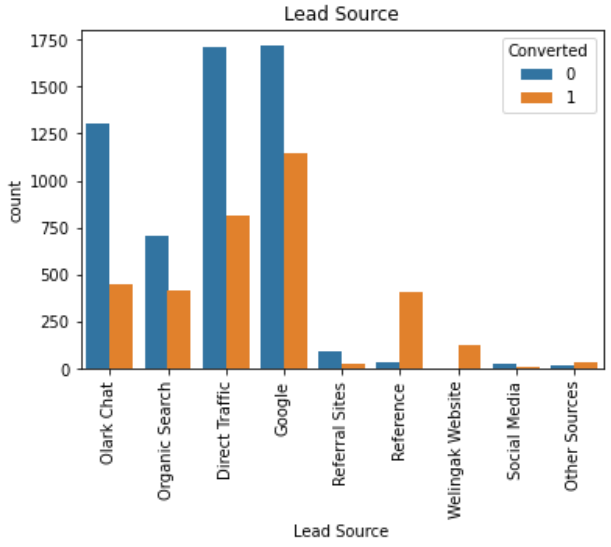
Through this process, some of the leads get converted while most do not.

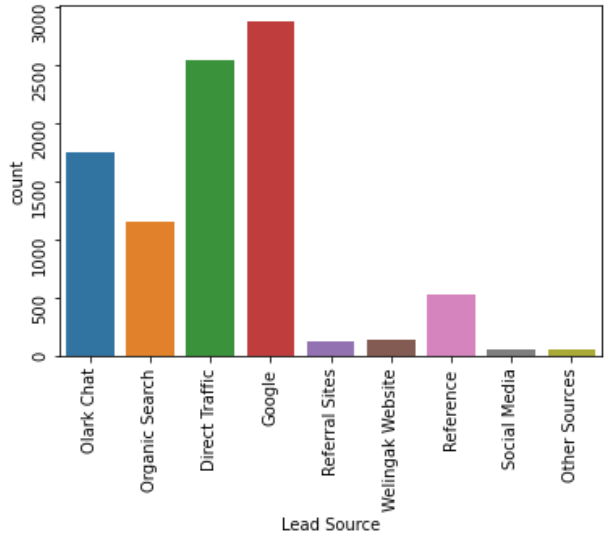
OBJECTIVE

To make this process more efficient, the company wishes to identify the most potential leads, also known as ‘Hot Leads’. The company requires to build a model wherein you need to assign a lead score to each of the leads

Approach


Import the data in python
Clean and visualize the data to understand it better
Treat the missing values, treatment is different for different columns
Preparing data for model building
Treating outliers by removing upper quartile and lower quartile
Building a logistic regression model
Scaling the data after train and test split
Evaluating the model on different measures and metrics
Measure the accuracy of the model and other performance metrics



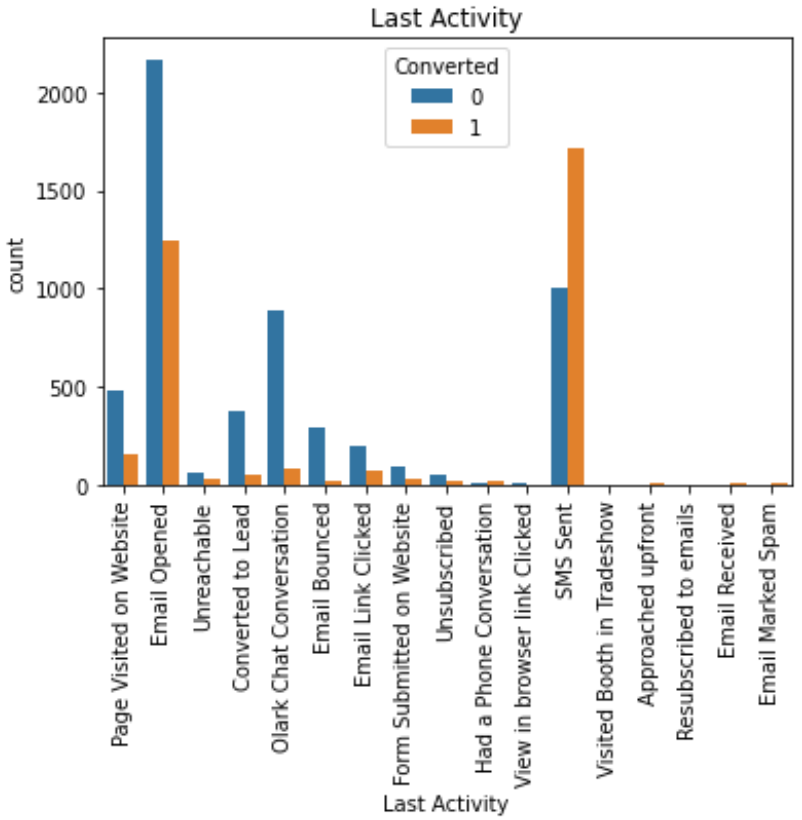


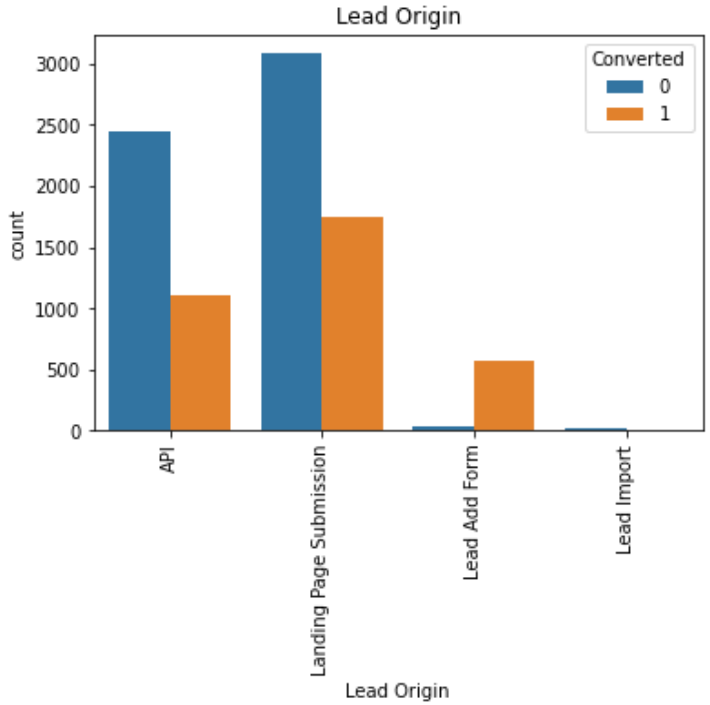
Univariant and Bivariant analysis

- Google searches have the highest conversion rate as compared to other options

  Direct traffic is next and then organic and Olark chat

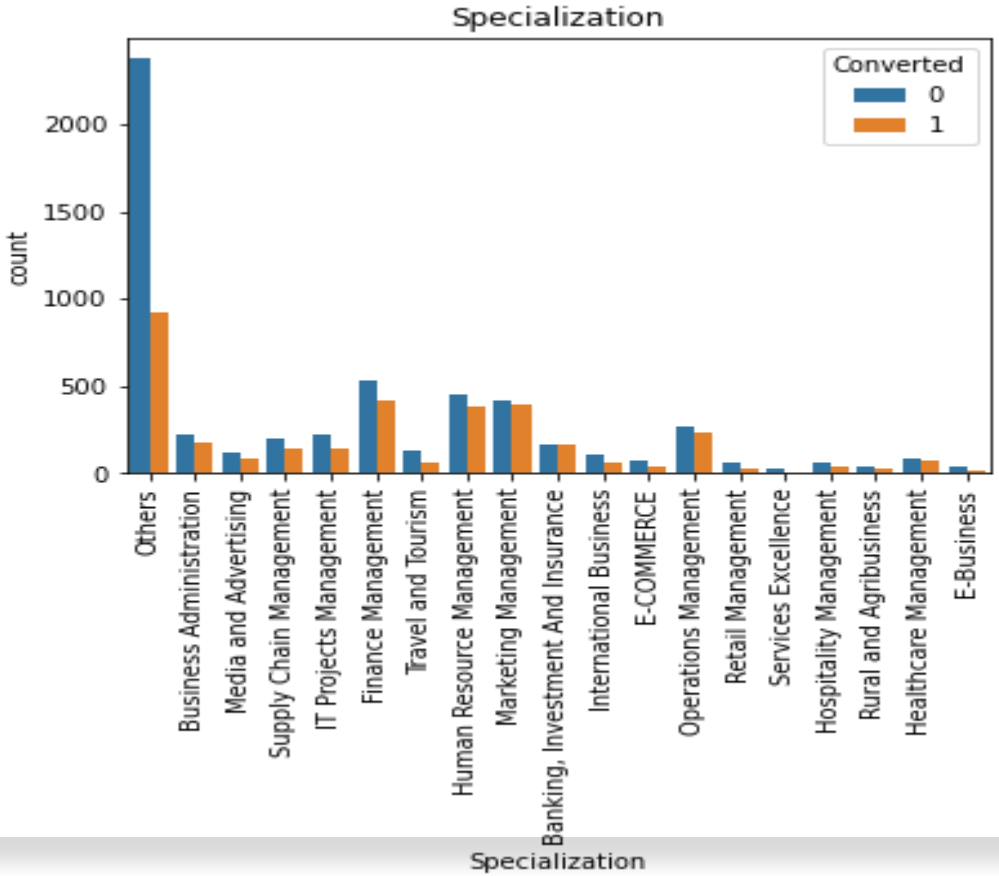
  Social media and reference sites fall behind

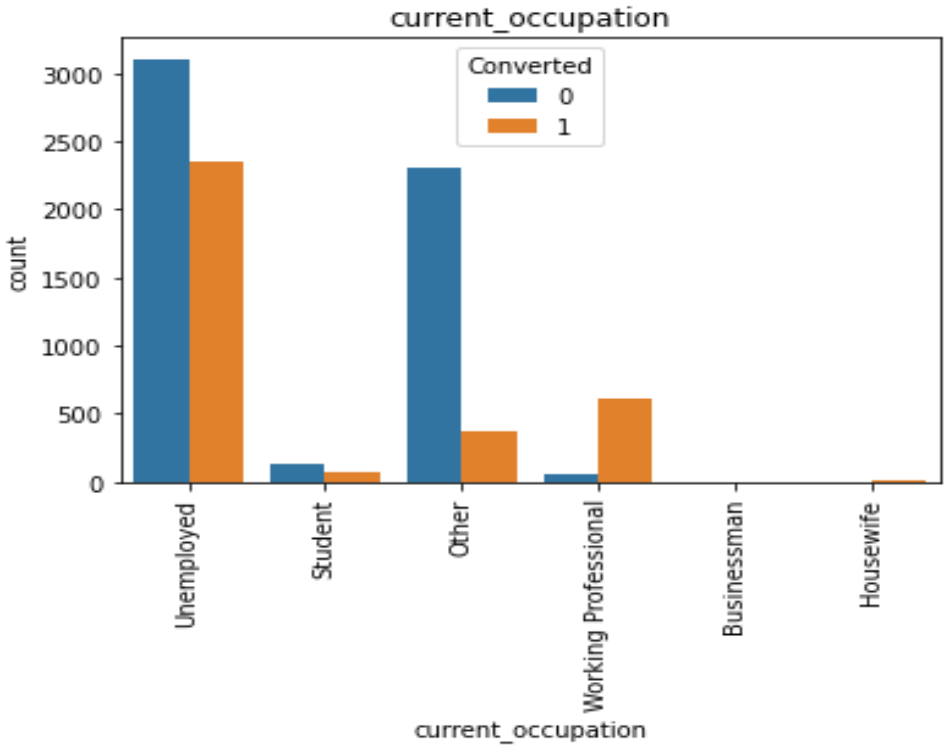




Last Activity and Lead Origin

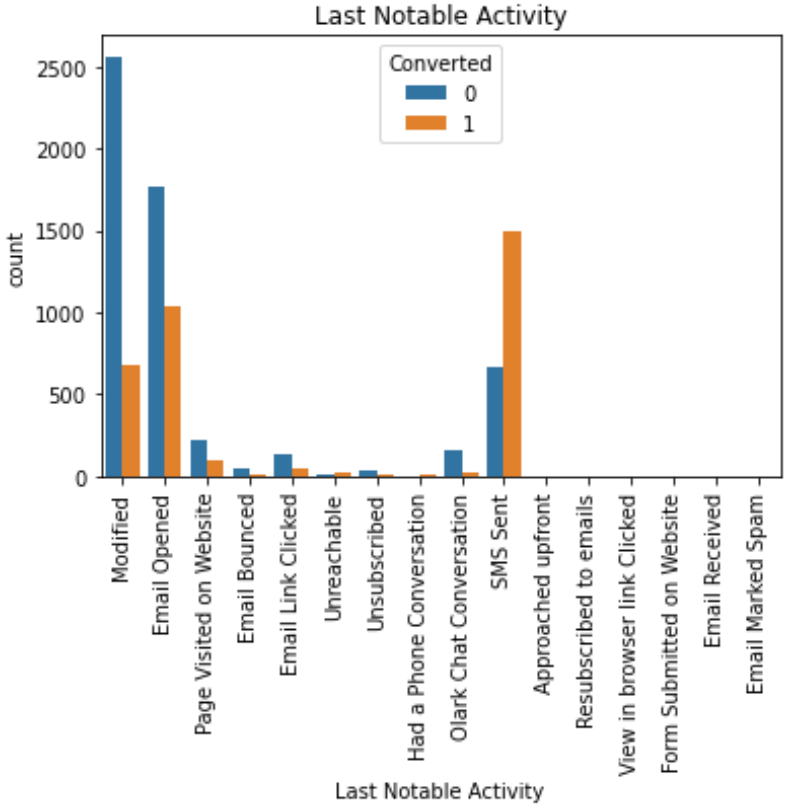
- Customers who communicated vis SMS have highest conversion rates
- Customers who end up submitting landing page and add form end up converting
- Lead import has less conversion rate

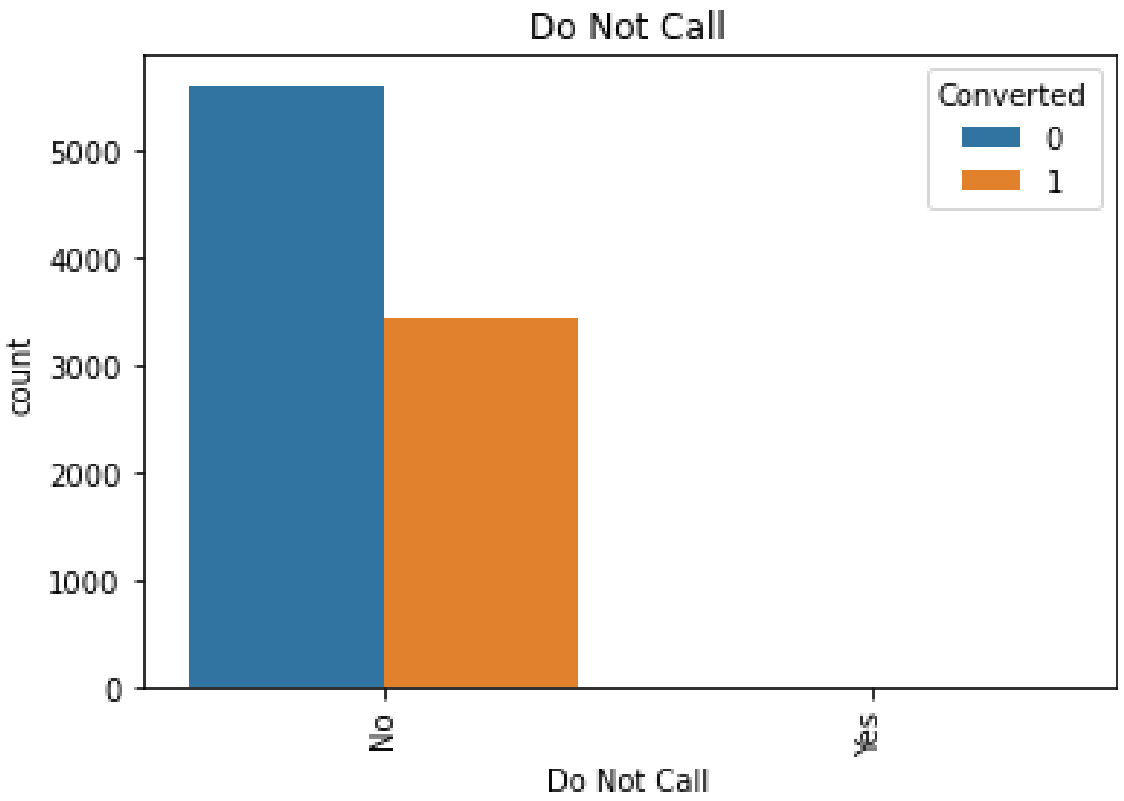




Specialization and Current Occupation

- Specialization of majority is not known so not a very clear picture, however
- Finance management , Marketing management specialization stream customers have high conversion rate
- Unemployed are more likely to convert to upskill themselves for the job market



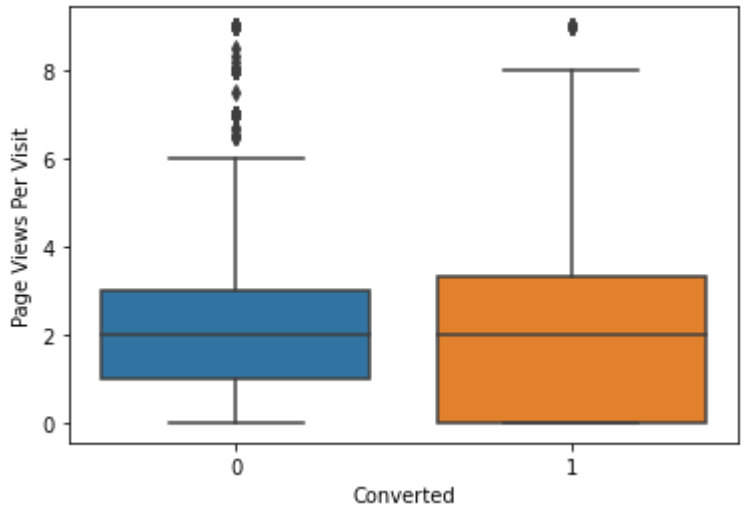


Last Notable Activity and Do not Call

- SMS sent as the last communication with the customers have high lead score conversion rate
- Most converted customers did not like to be called

Steps followed during data cleaning

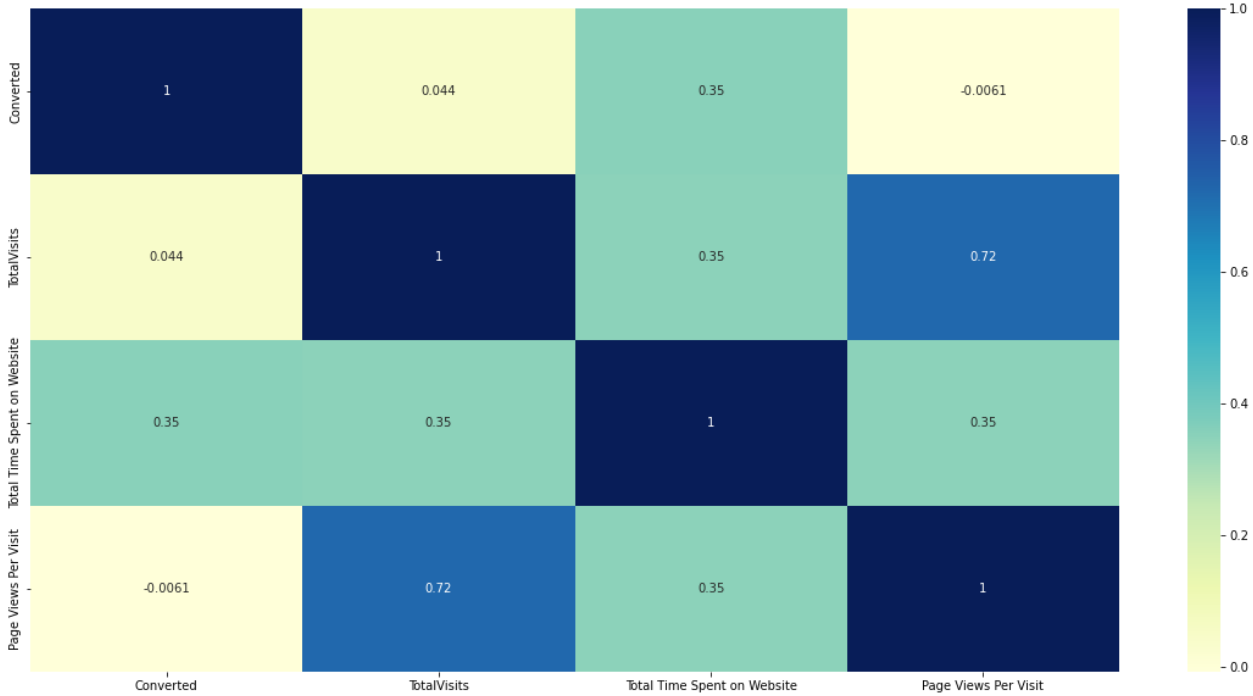
- Duplicate rows were checked
- % Null values for each column were calculated
- Those columns having more than 45% null values were dropped
- Lead Source null values were saved as other sources
- Country column was reassigned as India, Other Countries, USA, UAE
- Specialization column null values were assigned as not known
- Now no of unique values were calculated for each categorical column
- The ones with only one unique value are dropped as they are not contributing to the model building
- Reason for course nan values reassigned as not known
- Id and Lead number column are dropped as they are not adding value
- Source of information column nan values are categorized as not known
- Various columns have Select options also they are converted to not known or others in some cases
- Remaining columns with less than 1% missing values the respective rows were dropped
- City and country columns were also taken care of by categorizing them to not known category



Mean is same for both the cases however large no of ppl who visited the page / no of views, get enrolled to the program

Outliers were treated by removing top and bottom 1 % of the data

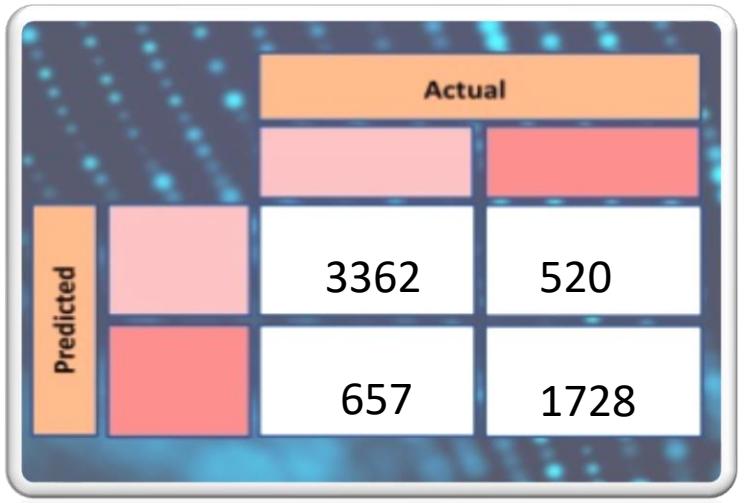
More time spent on website leads to the Conversion of customers

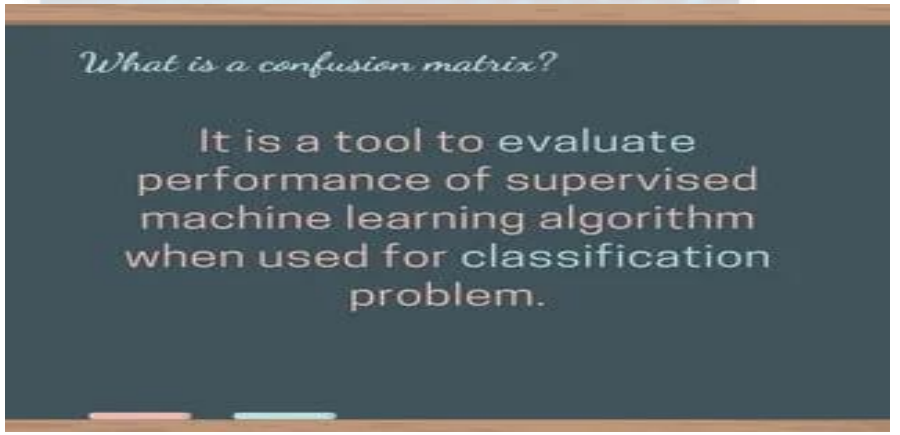


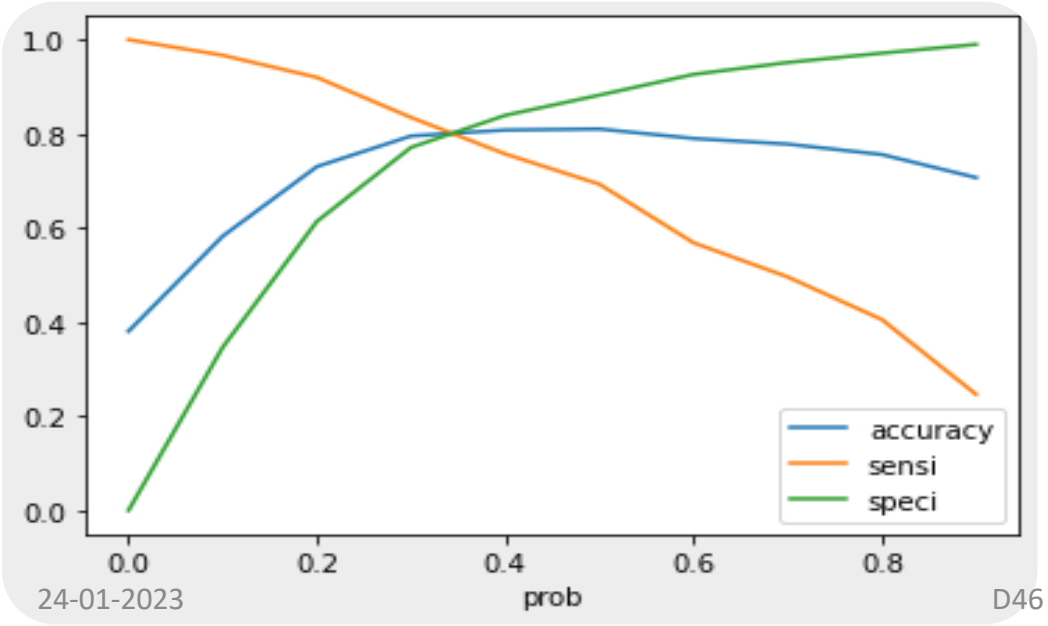
Model Building

Model 1 at threshold of 0.4

- Accuracy: 81.22%
- Sensitivity: 71.39%
- Specificity: 83.60%
- FPR : 11.40%

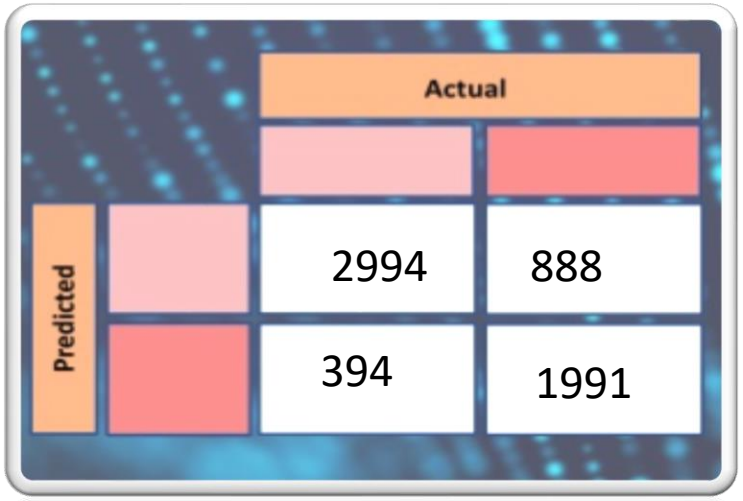






Threshold choosen as 0.3

Model 2



- Accuracy: 79.54%%
- Sensitivity: 72.45%
- Specificity: 86.60%
- FPR : 13.40%
- Positive predictive value: 76.87%
- Negative predictive value: 83.65%

Precision : 69.16%
Recall: 83.48%

Test dataframe model performanceevaluation matrix

- Accuracy: 79.04%
- Sensitivity: 81.09%
- Specificity: 78.82%
- Precision score: 67.67%
- Recall: 84.75%

- Positive predictive value: 76.87%
- Negative predictive value: 83.65%

Conclusion

The top three variables are:

- Total time spent on website
  i. The total time spent by the customer on the website.
  ii. The more the time spent means more likely to convert.

- Lead Origin
  i. The origin identifier with which the customer was identified to be a lead. Includes API, Landing Page Submission, etc.
  ii. If the lead fills out a lead add form more likely to convert
- Lead source
  i. The source of the lead. Includes Google, Organic Search, Olark Chat, etc.
  ii. Of these ppl contacting through Olark chat are more likely to convert

Conclusion

Top dummy variables are:

- Lead Origin- Lead Add form : Customers filling out add forms are more likely
- Lead Source – Olark Chat : Customers engaging in an Olark chat window are more likely to convert
- Last activity – SMS sent: If customer has sent an SMS, more likely to convert

Suggestions

- First and foremost, we should arrange the customers based on the amount of time spent by them on the website. The more the time spent, strong likelihood of them converting to avail the course, so continuous follow up and messages should be sent.
- Targeting the customers who engage on the chat or have come through Welingak Website. Those customers who have last communicated through SMS are also most likely to get converted so need to be followed up regressively. The queries resolution should be quick via SMS and a quick call-in case some more information needs to be communicated.
- The students at some school, and those customers who have marked do not contact via email, are very unlikely to be converted. So, resources need not be wasted on following up, however basis information asked should be communicated promptly.

The customers spending long time, messaging via SMS, filling out the ADD form and chatting on Olark chat are most likely to get enrolled to the course. So, the website experience should be good and smooth.

Suggestions for better customer enrolment rate

Make groups and follow up based on the categories of customers like strong leads, not so strong and weak leads. Make email draft to be sent to these categories

Keep the draft emails short and engaging referring to the alumni achievements or where are they placed.

Catch the attention by keeping some free engaging sessions on the subject matter.

Consider using an informal and personal touch by sighting the same case scenario.

Never assume whether its worth deleting cold leads, let the customer tell that themselves.

Keep prerecorded voicemails and email templates for various scenarios that can be sent with little tweak.