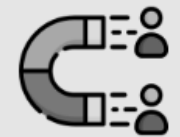
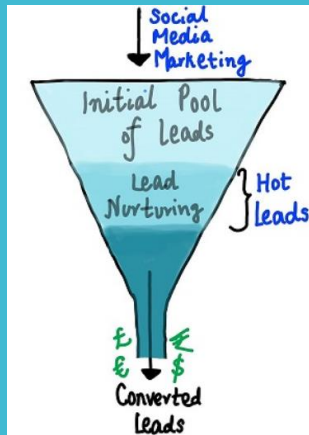


Lead Score Case Study

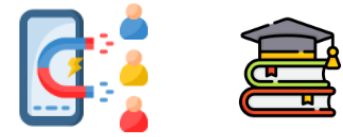
BY : AMANPREET KAUR WALIA, AMRITA KHARE, ANBU RAMASAMY



PROBLEM STATEMENT



- X Education(an education company) sells online courses to industry professionals. People who are interested in the courses land on their website and browse for courses
- When they fill up form with their email id or phone number, they are said to be a lead. The company also gets leads through past referrals.
- On acquired leads, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted . The typical lead **conversion rate** at X education is around 30%.
- For more efficiency, the company wishes to identify the most potential leads, known as '**Hot 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



- X education wants to know hot leads
- Need to build model to identify most promising leads with lead conversion rate of 80% or more
- Deployment of model

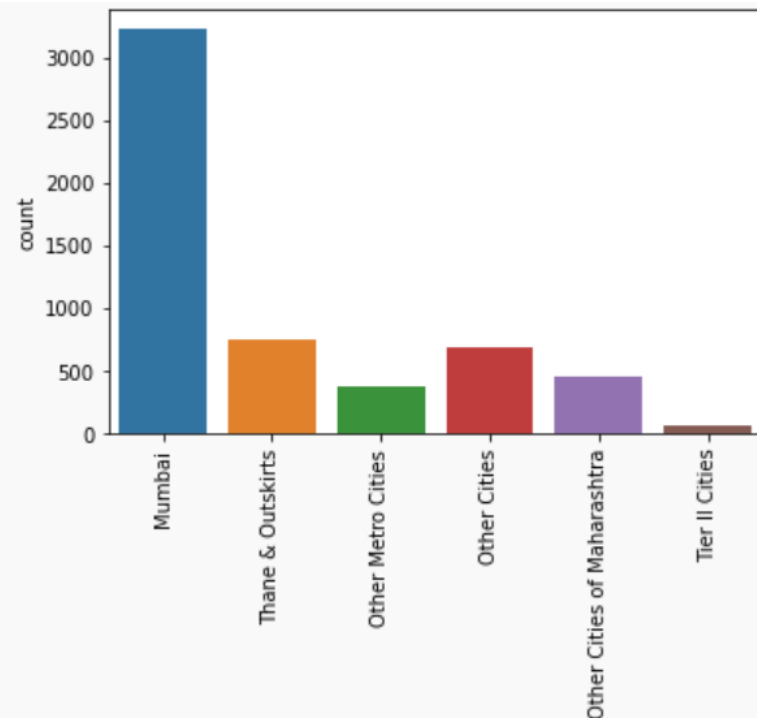


Methodology

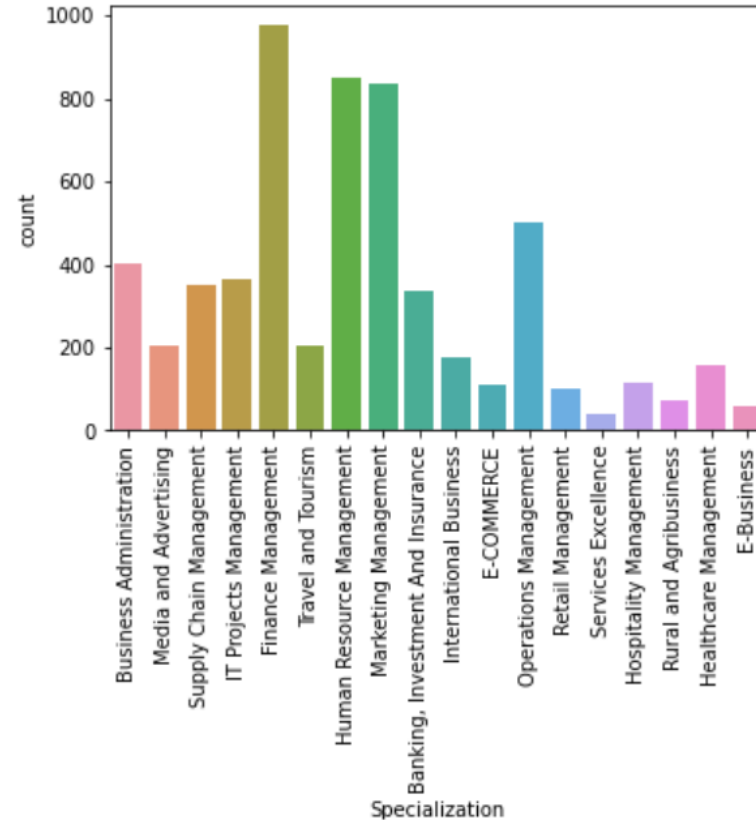
- Data cleaning and data manipulation
- check and handle duplicate data
- Check and handle NA values and missing values.
- Drop columns, if it contains large amount of missing values and not useful for the analysis.
- Imputation of the values, if necessary.
- Check and handle outliers in data.

Analysis

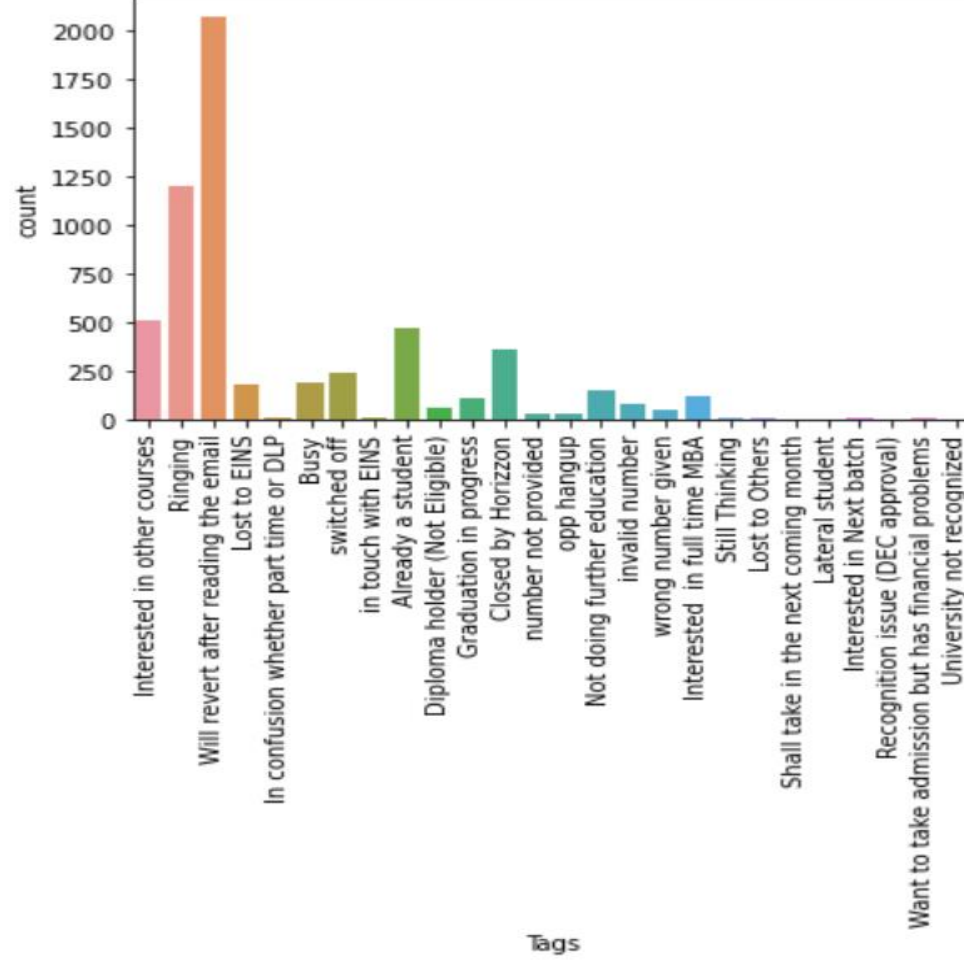
For cities



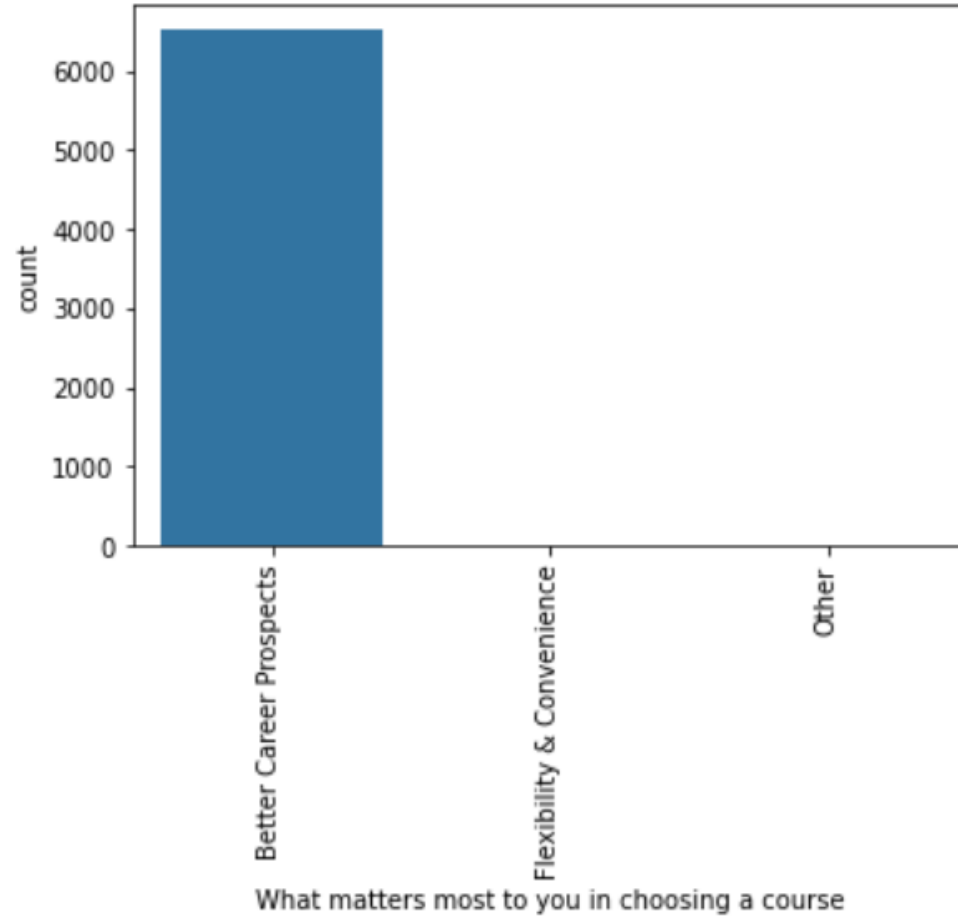
Specialization



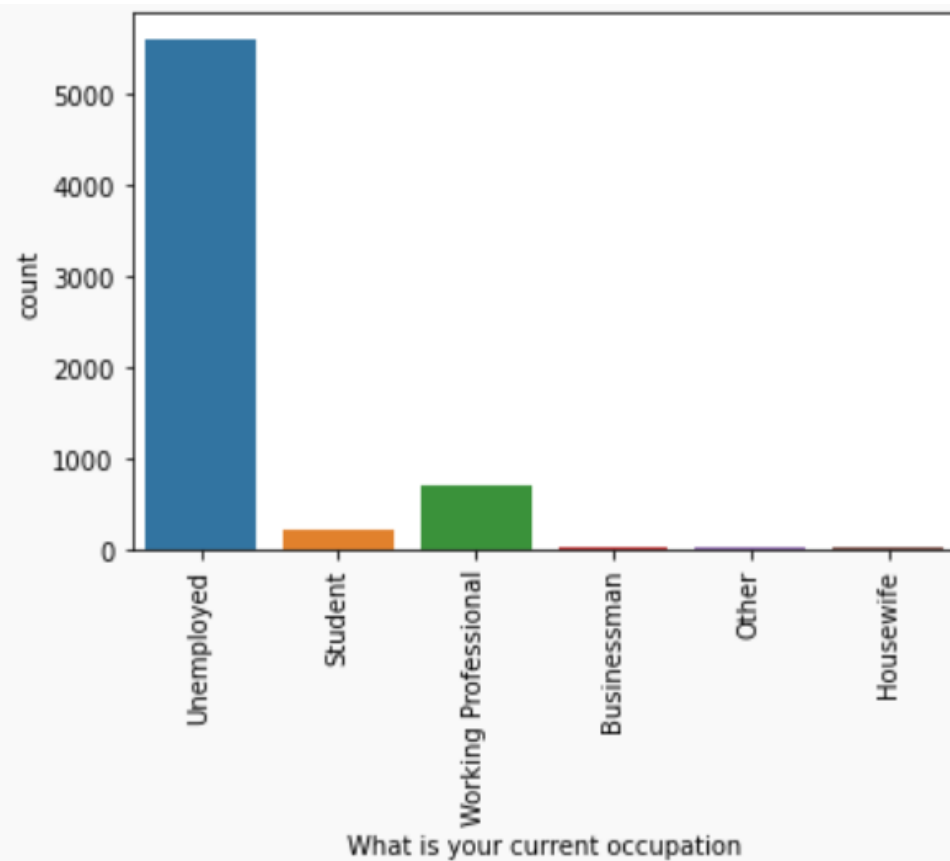
Tags



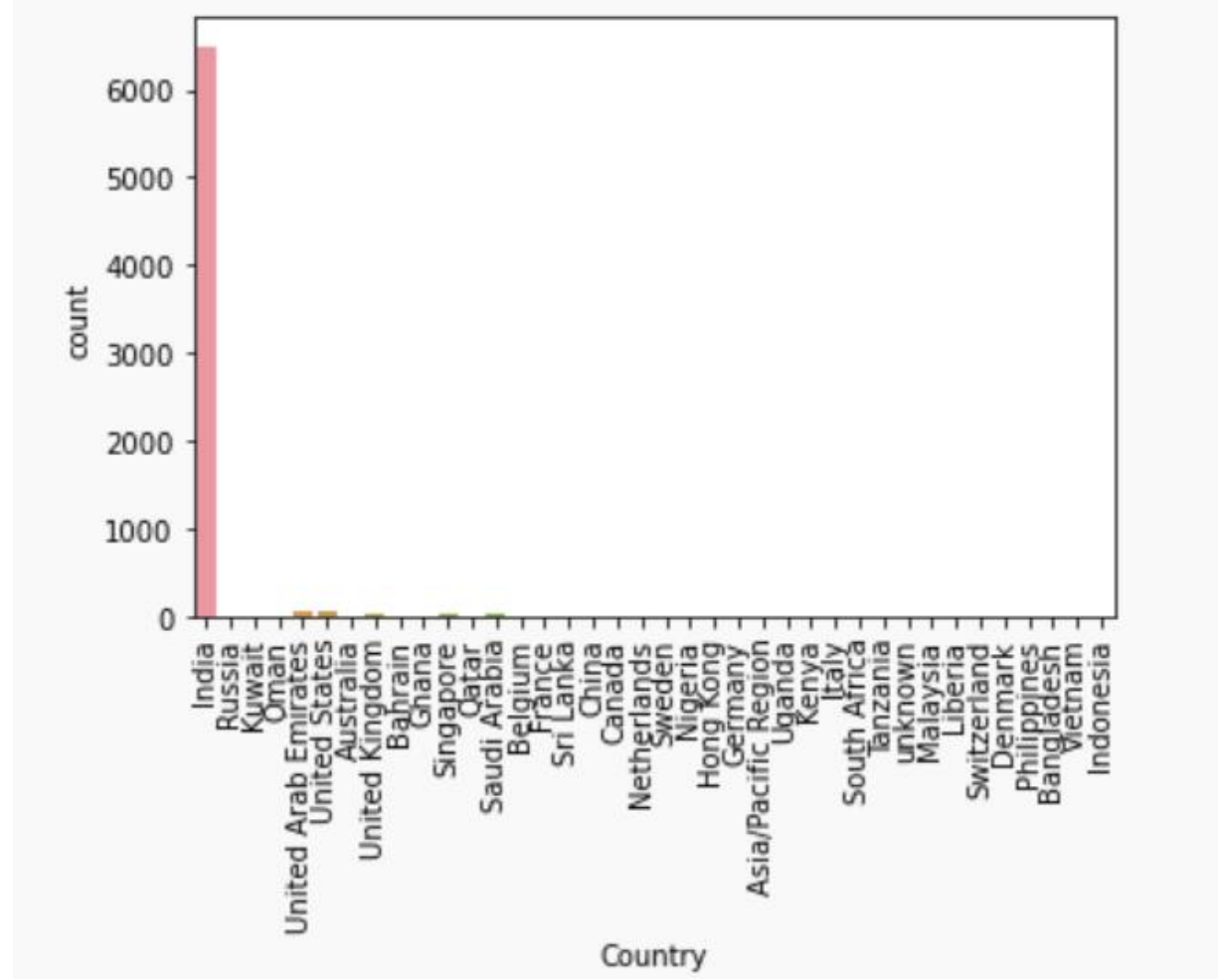
What matters



Current Occupation

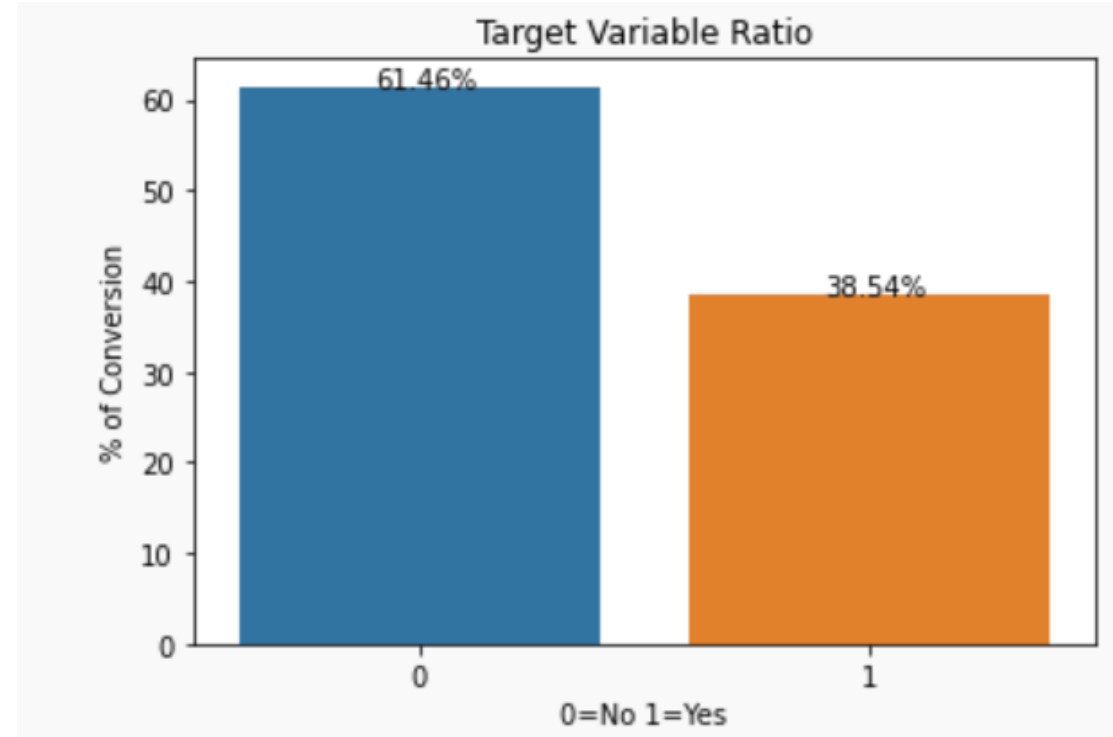


Checking country

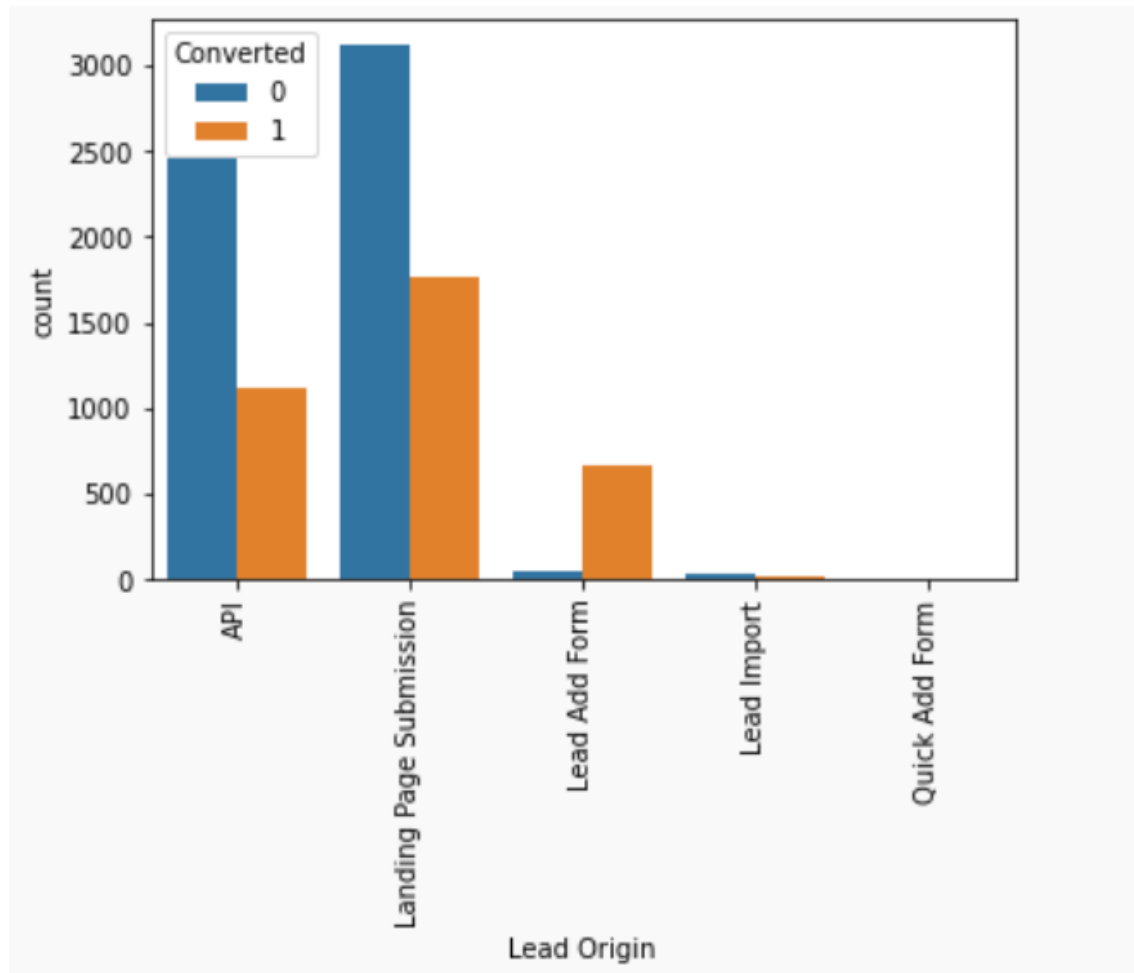


Data visualization

Converted
data



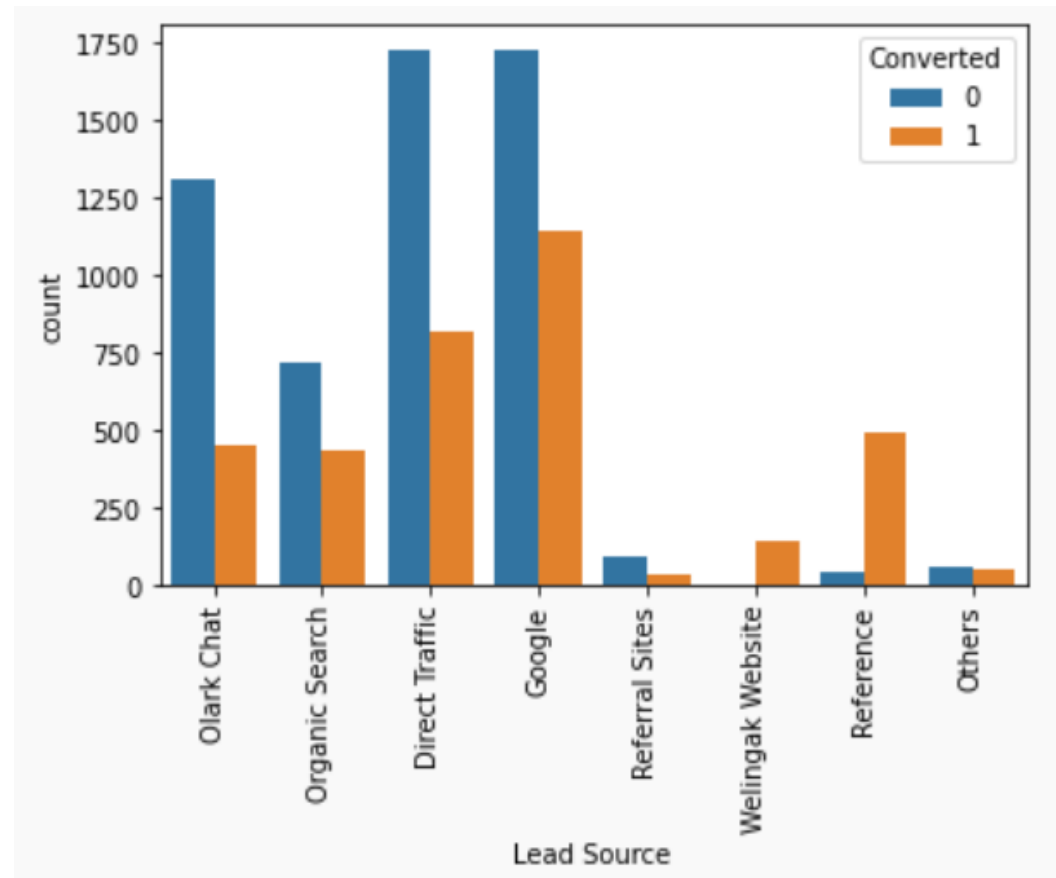
Lead Origin



Inference from Lead Origin

- API and Landing Page Submission have around 35-40% conversion rate but count of lead originated from them are considerable.
- Lead Add Form has more than 90% conversion rate but count of lead are not very high.
- Lead Import and Quick Add From are very less in count.

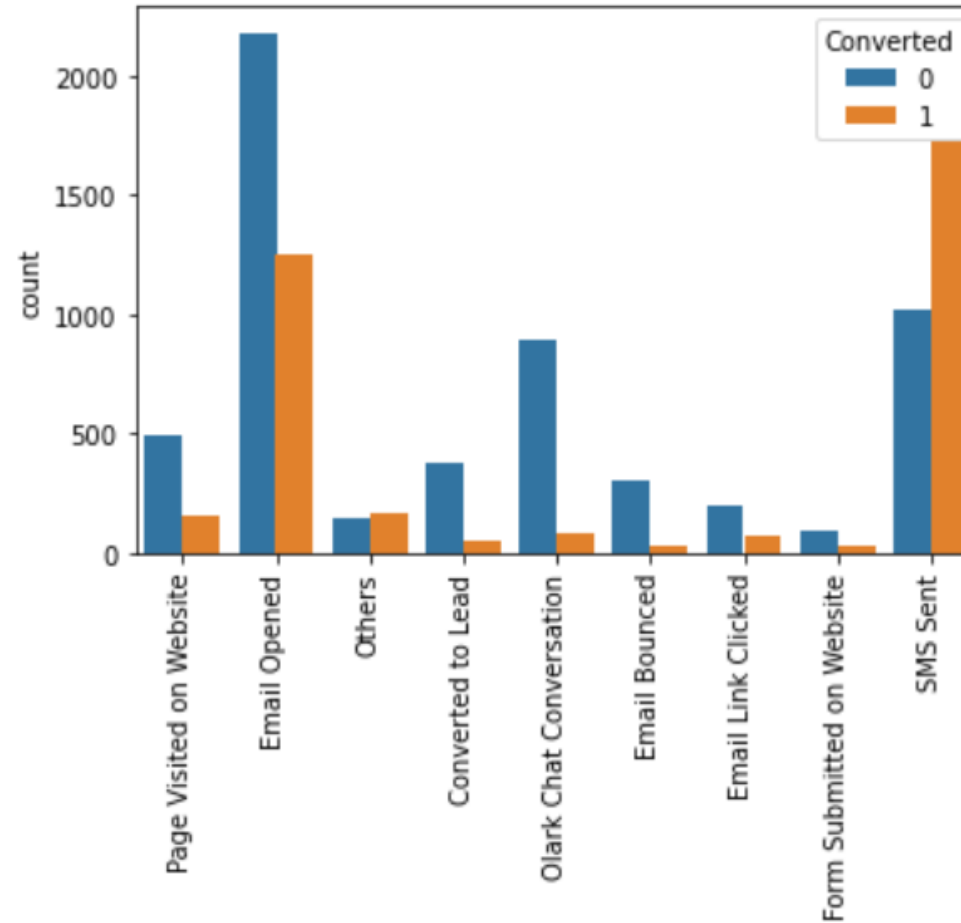
Lead Source



Inference from Lead Source

- Direct Traffic and Google are generating high number of leads; should be focused for their conversion
- Welingak and Reference are having more conversion rate hence should be focused to generate more leads
- leads from Olark Chat and Organic Search should be focused for conversion

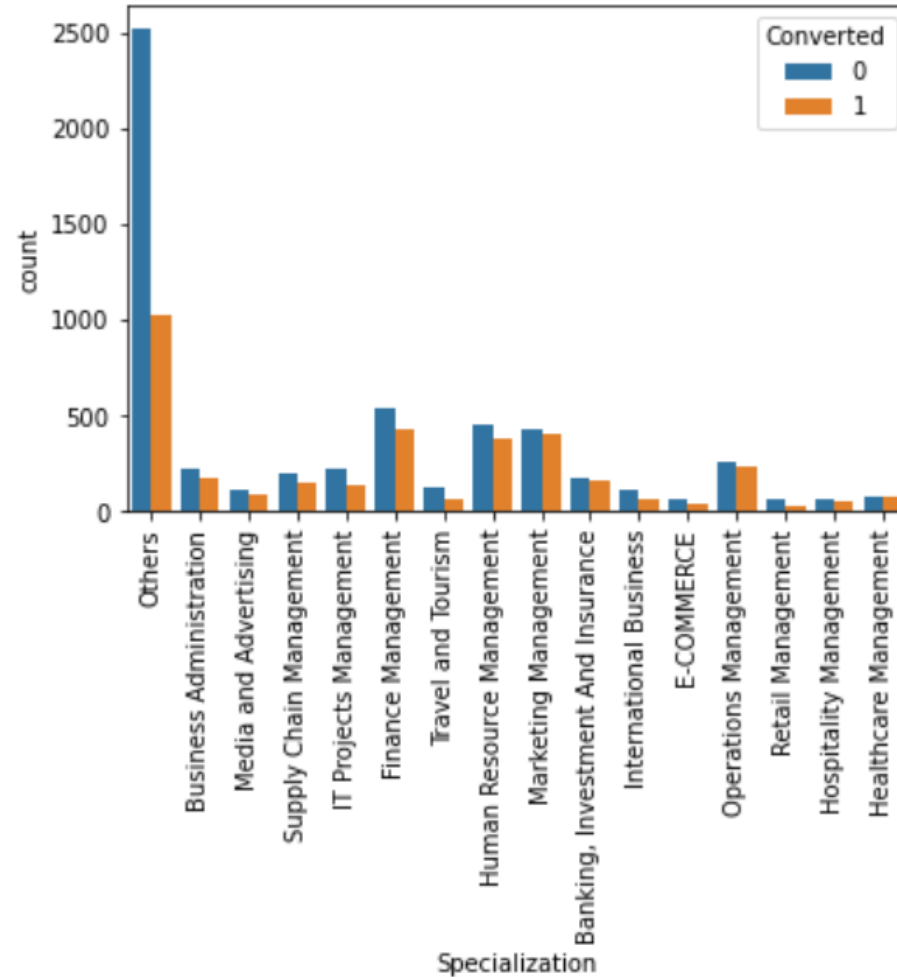
Last Activity



Inference from Last Activity

- Most of the leads have Email Opened and SMS Sent. More focus should be on leads opening email.
- conversion rate for SMS sent is very high and hence they should be targeted for lead conversion

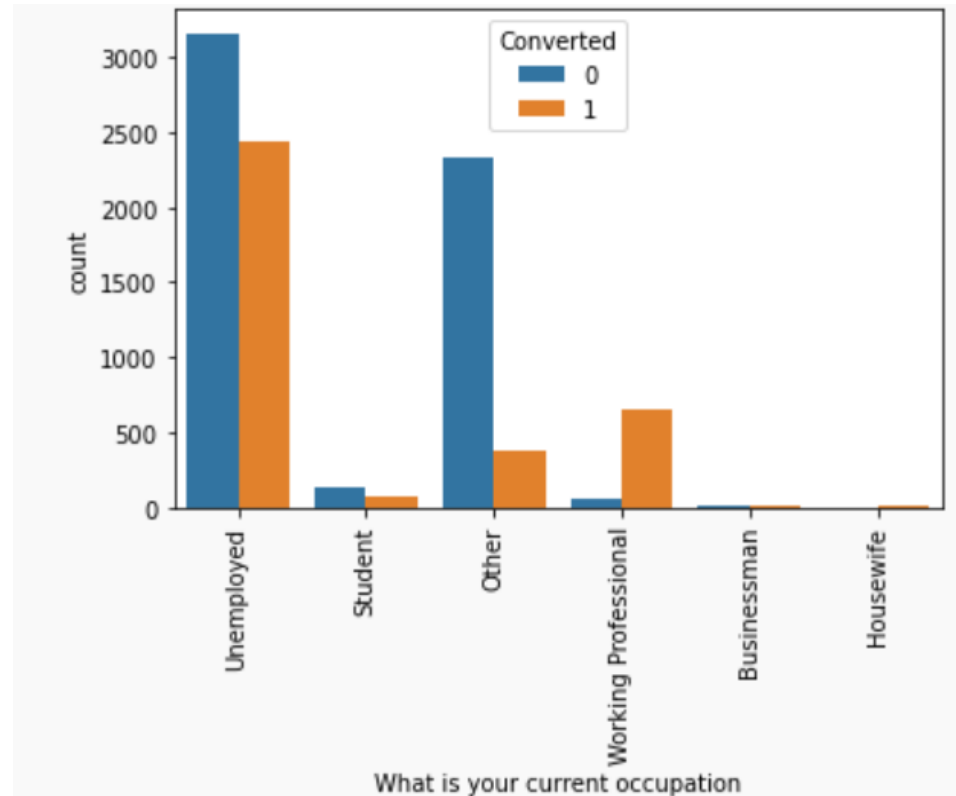
Specialization



Inference from Specialization

- focus should be on those various specializations with less leads but high conversion rate.
- Most of the leads are from 'Others'(Not specified) specialization but the conversion rate is poor

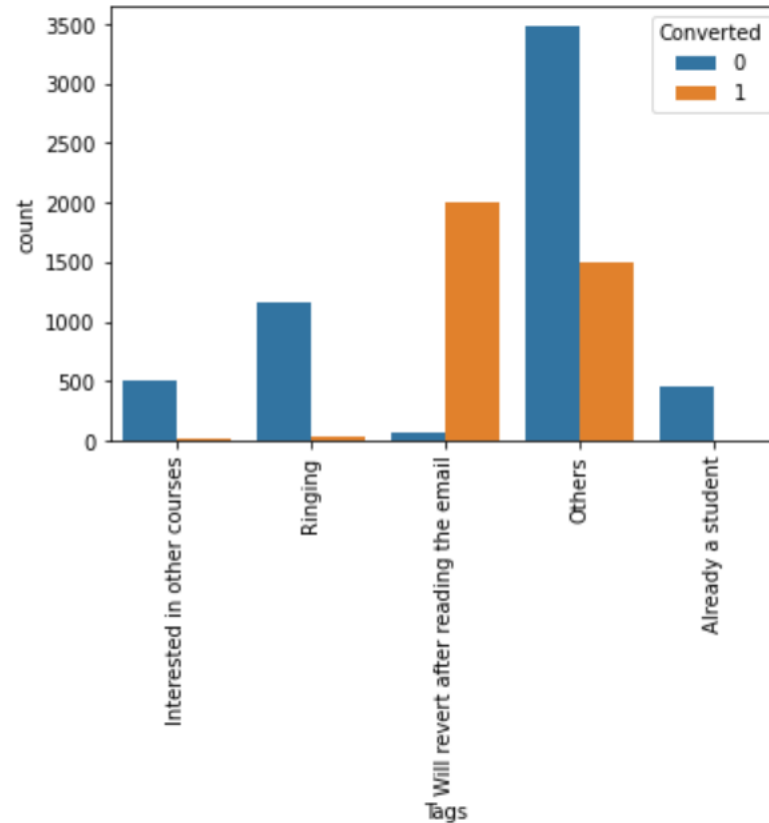
Current occupation



Inference from What is your current occupation

- Working professionals have high conversion. focus should be on them for more leads
- Most of the leads are from Unemployed section, focus should be increasing conversion from them

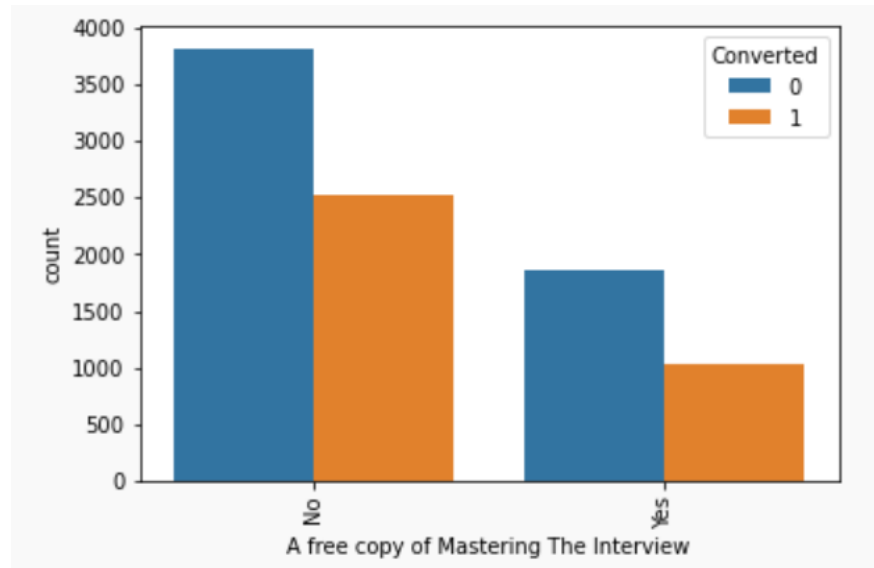
Tags



Inference from Tags

- 'Will Revert after reading the email' have high conversion. focus should be on them for more leads
- Most of the leads are from 'Others'(Unspecified) section, focus should be increasing conversion from them

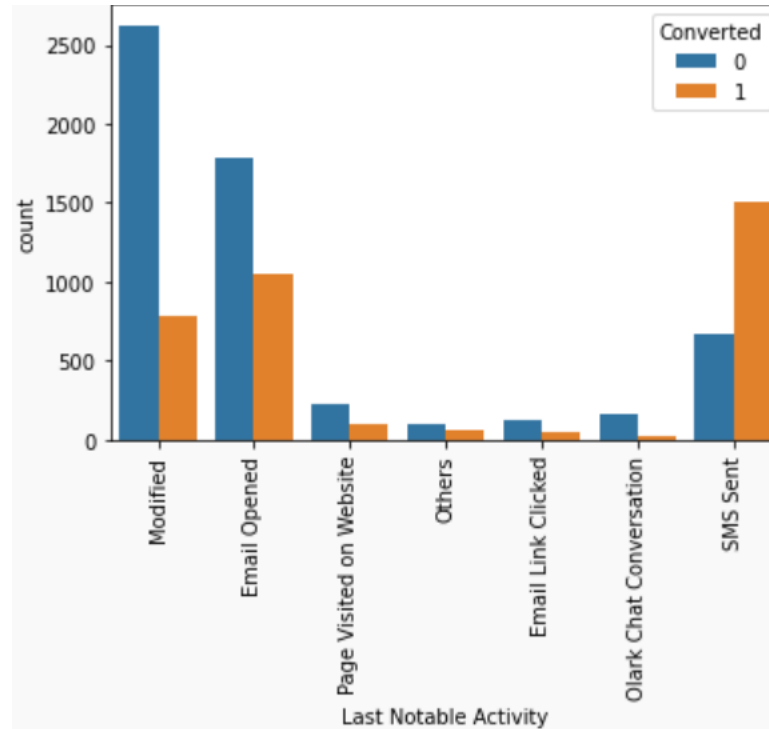
Mastering the Interview



Inference from A free copy of Mastering The Interview

- More leads are from those who do not ask for free copy of Mastering Interviews. Can be focused for conversion

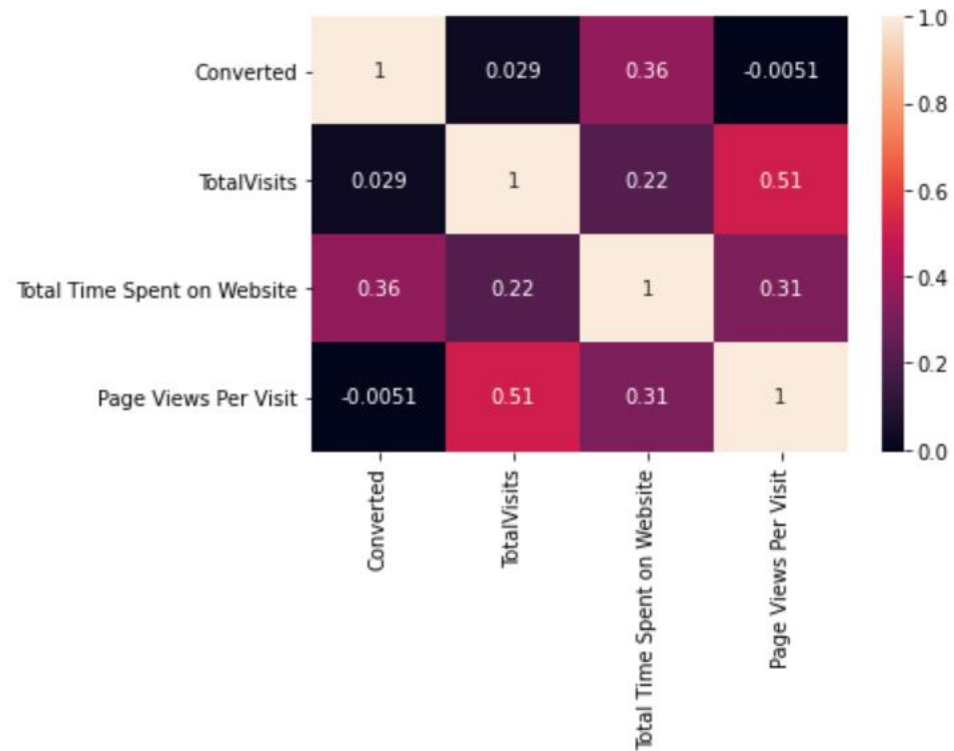
Last Notable Activity



Inference from Last Notable Activity

- More leads are from those who have modified their account or opened email. focus should be on their conversion
- SMS sent have high conversion

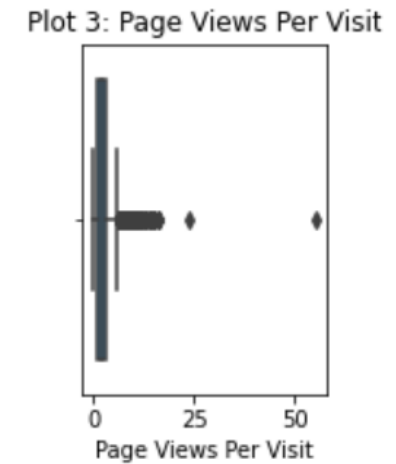
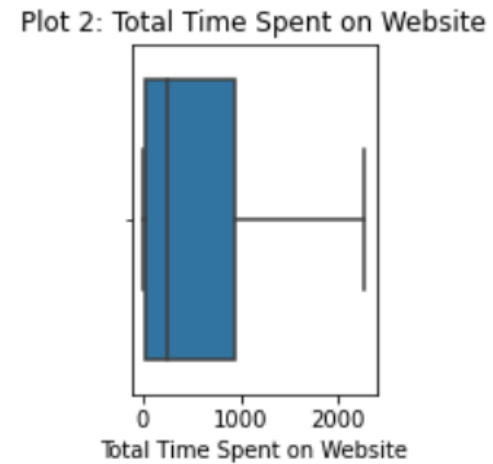
Heat map



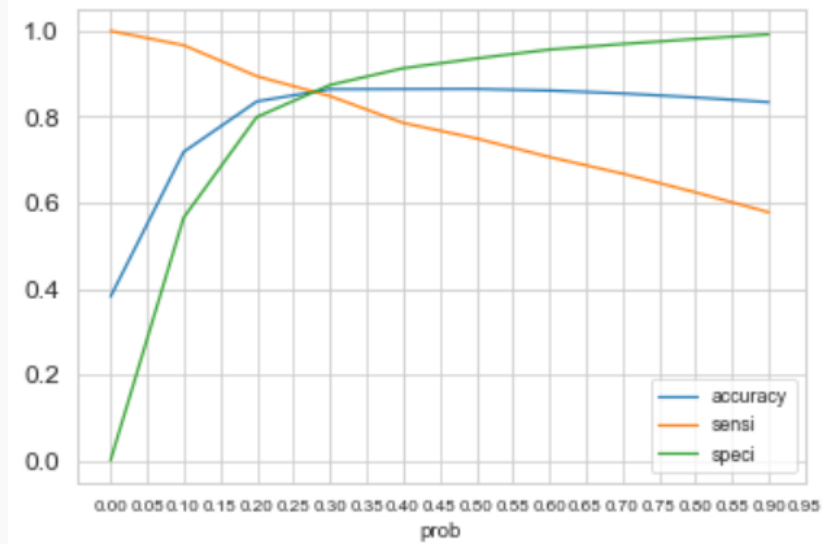
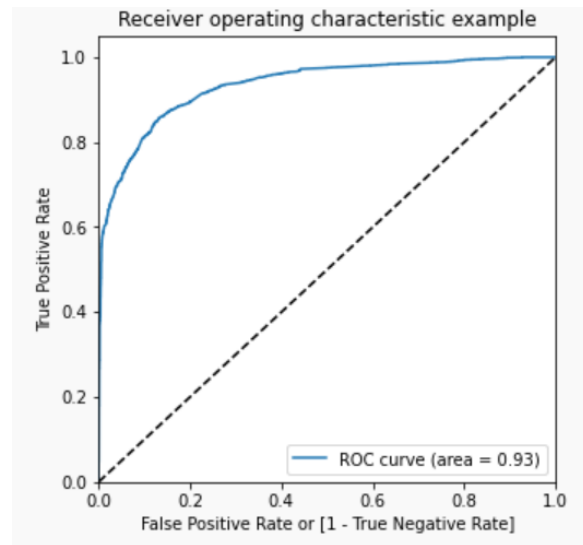
Inference from above heat map

- Strong correlation between Total Visits and Page Views Per Visits
- Converted has good correlation with Total time spent on Website

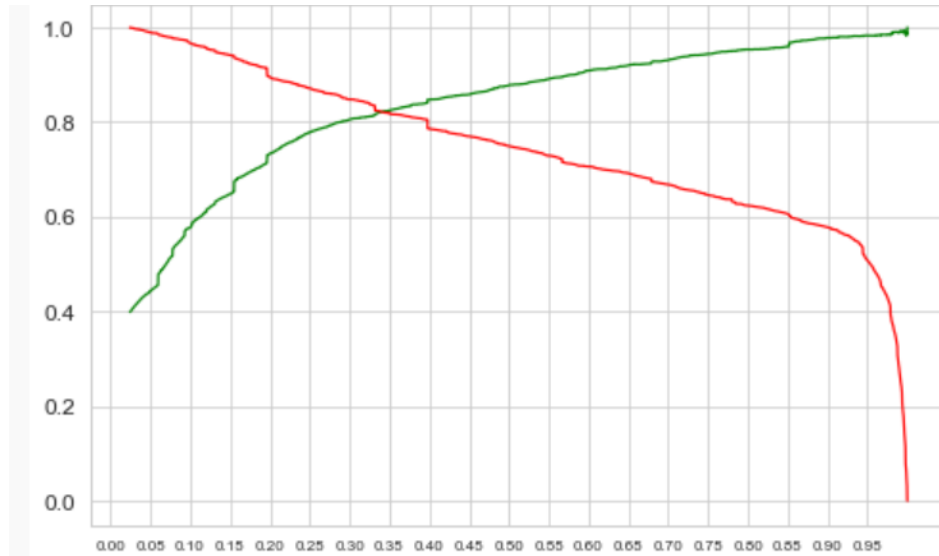
Check outliers



ROC Curve



Model Evaluation



Model Analysis



Analysis	Percentage
Sensitivity of our logistic regression model:	80%
Specificity of our logistic regression model:	81%
Overall accuracy on Test set :	80.34%



It was found out that the variables that mattered the most in the potential buyers are :

- The Total Time Spent on Website.
- Total number of Visits.
- When The Lead source was :a. Olarkchat b. Wellingak Website
- When the Last Activity was :a. SMS b. OlarkChat Conversation
- When the lead origin is Lead add Form.

When the current occupation was : Working Professionals. Student. Unemployed. Other

Keeping The above mentioned points in mind the X education can increase all the potential buyers to change their mind and buy their courses.

Conclusion