X Education - Lead Scoring Case Study

Making a model for identifying hot leads to increase conversion rate for X Education Company

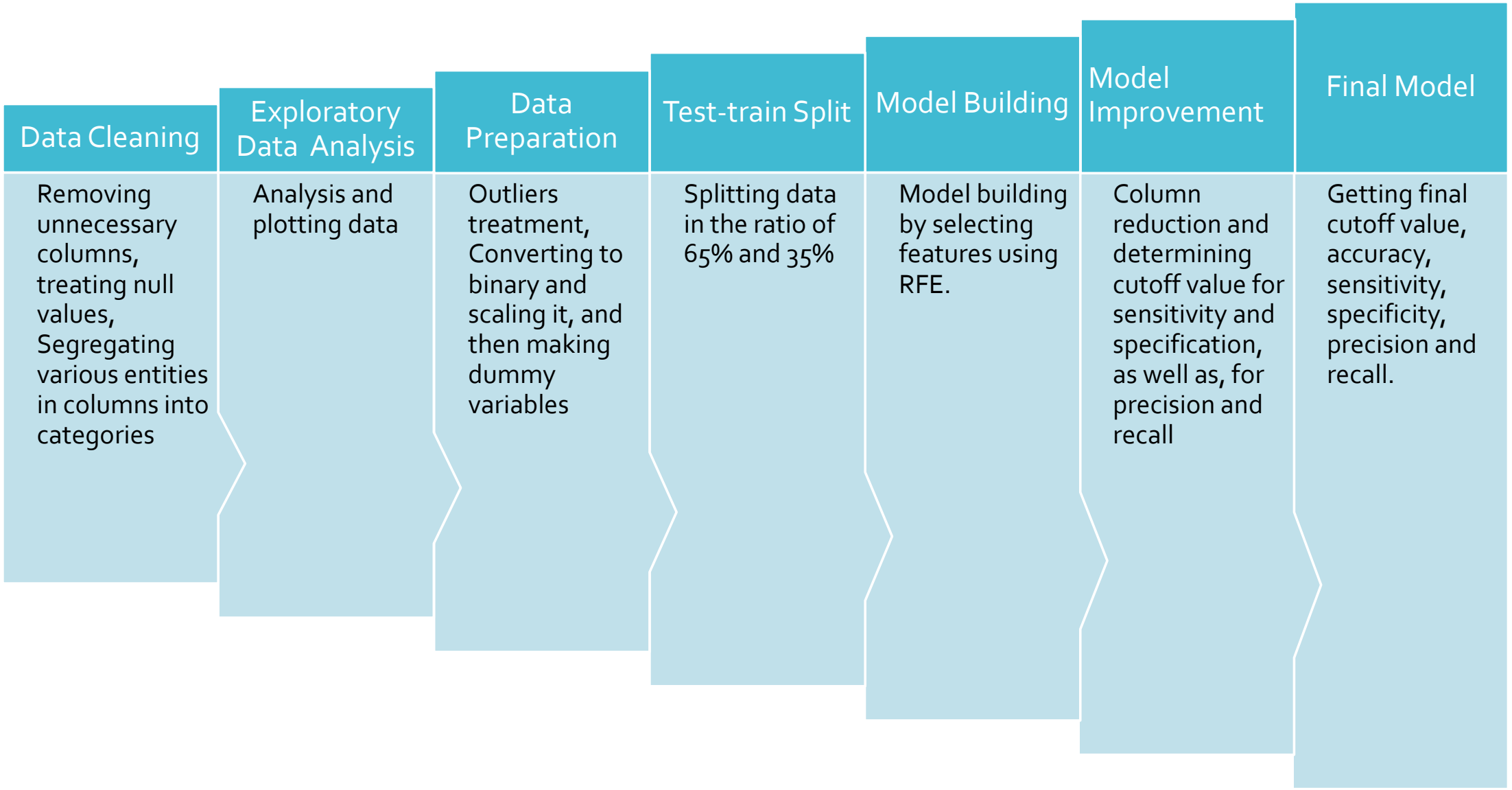
PROBLEM STATEMENT

- X Education is an online education company that offers courses to industry professionals.
- They market their courses on various websites, and interested people visit their website to browse and fill out forms.
- Those who provide their contact information are considered leads. The company gets leads through past referrals as well.
- The sales team makes calls and sends emails to convert the leads, but only around 30% of leads get converted.
- X Education aims to identify the most promising leads, also known as 'Hot Leads,' to improve their lead conversion rate. By focusing on these leads, the sales team can communicate more effectively and increase conversions
- X Education has asked to develop a model that assigns a lead score to each lead, indicating the likelihood of conversion. The CEO has set a target lead conversion rate of 80%.

SOLULTION

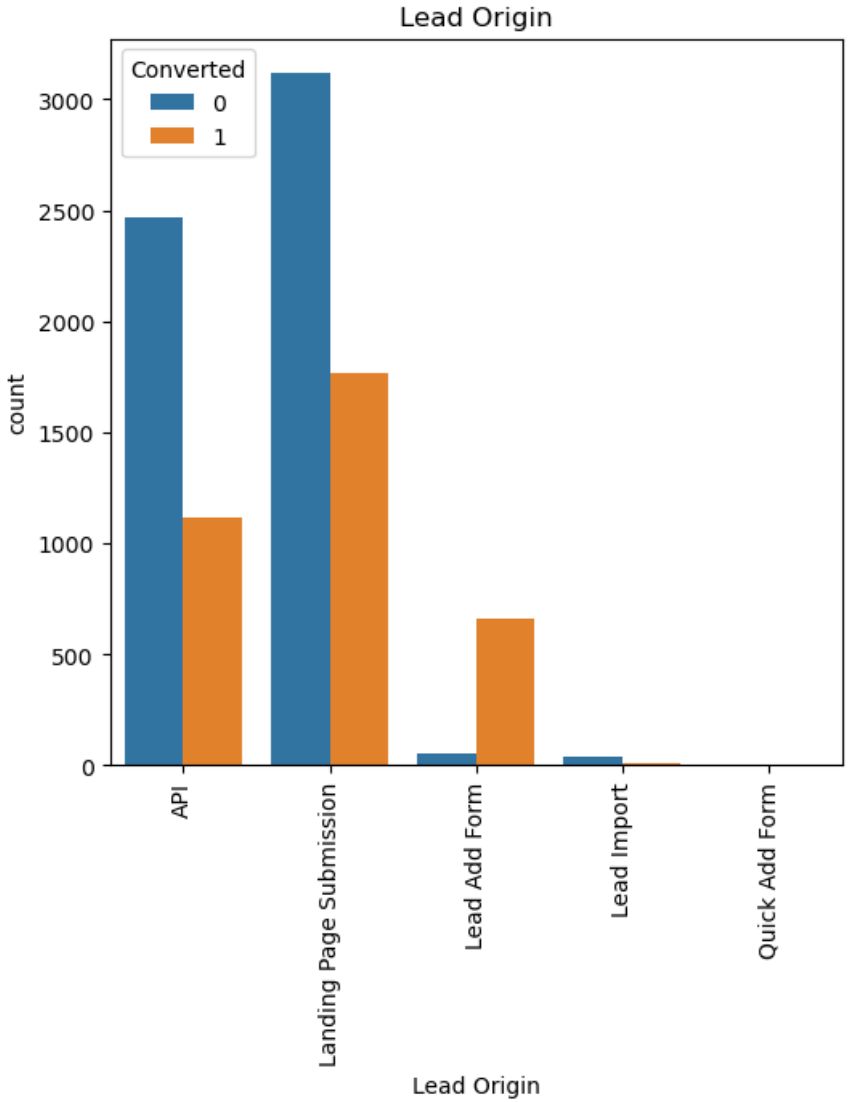
- The solution to problem statement given is to create model which can identify high potential lead or ‘hot leads’
- We have to make a model and workout a cutoff based on scores obtained by various applicants, which will help us identify a potential lead(Scores would range from 0 to 1)
- We will be aiming for good accuracy as we need conversion rate of about 80%

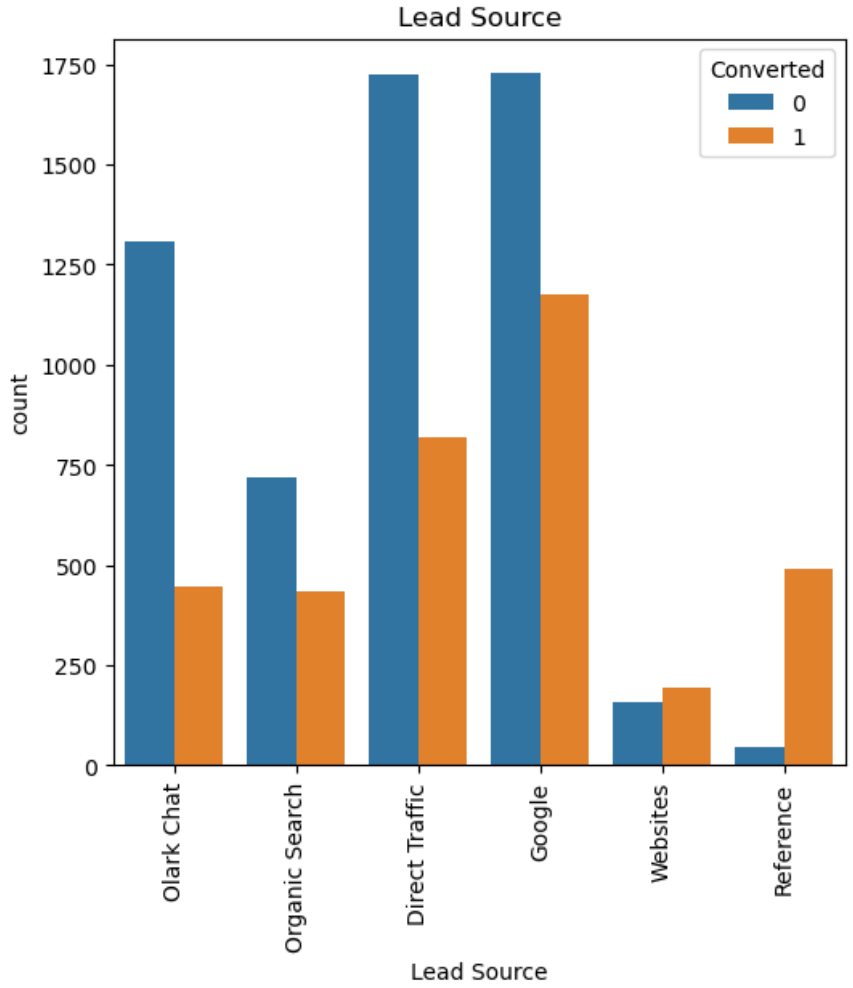
PROCESS FOR MODEL MAKING


Data Cleaning
Removing unnecessary columns, treating null values, Segregating various entities in columns into categories
Exploratory Data Analysis
Analysis and plotting data
Data Preparation
Outliers treatment, Converting to binary and scaling it, and then making dummy variables
Test-train Split
Splitting data in the ratio of 65% and 35%
Model Building
Model building by selecting features using RFE.
Model Improvement
Column reduction and determining cutoff value for sensitivity and specification, as well as, for precision and recall
Final Model
Getting final cutoff value, accuracy, sensitivity, specificity, precision and recall.

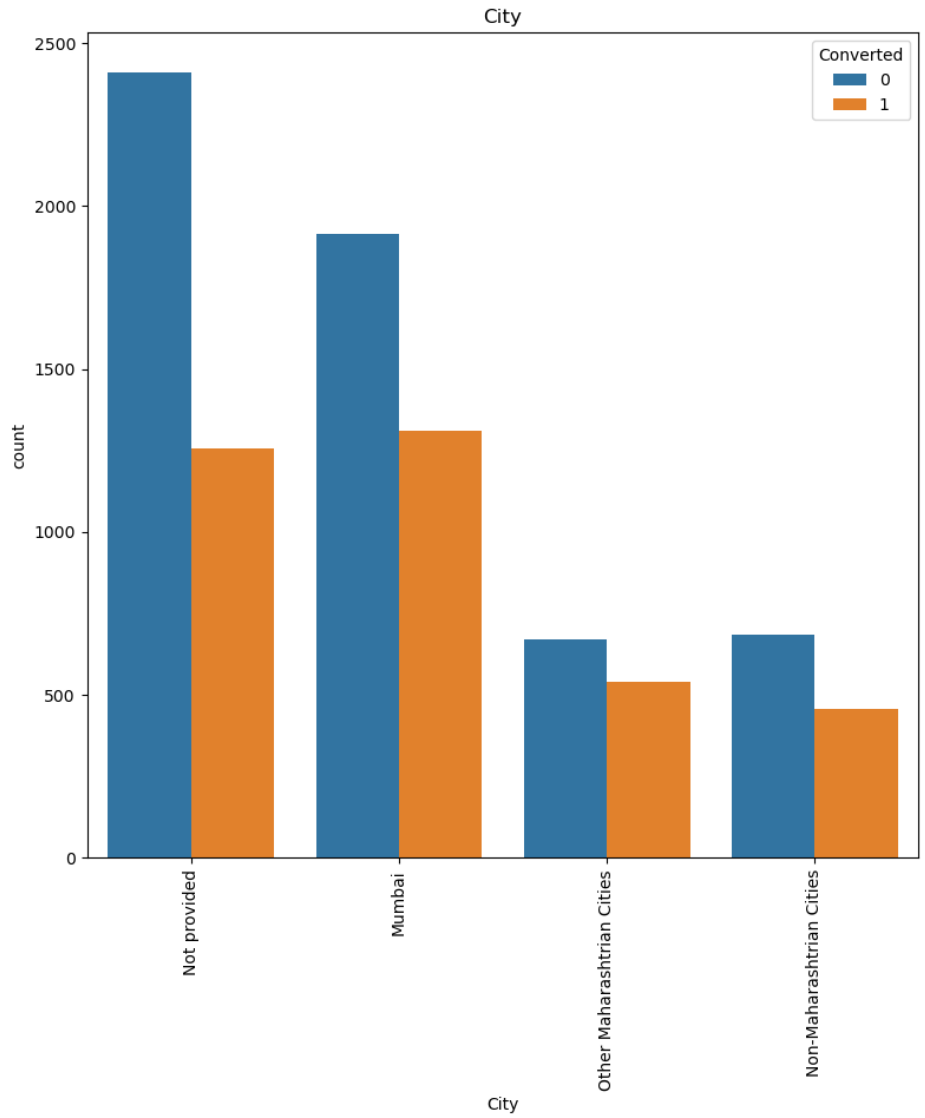
Visualizing data with graph

Graph comparing proportion of various leads which converted(orange) and didn't convert(blue) into potential customer from Lead origin and Lead Source column

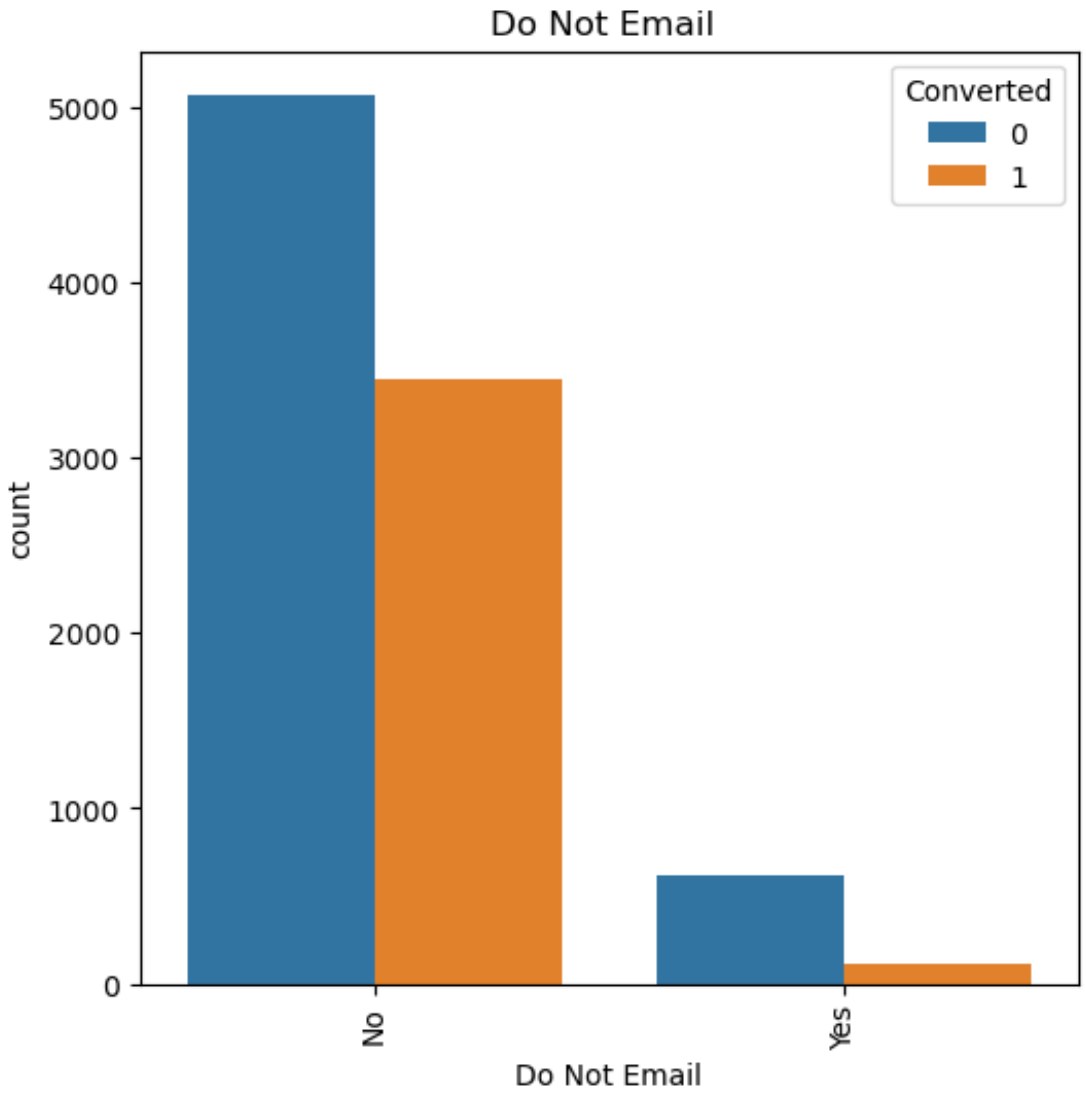




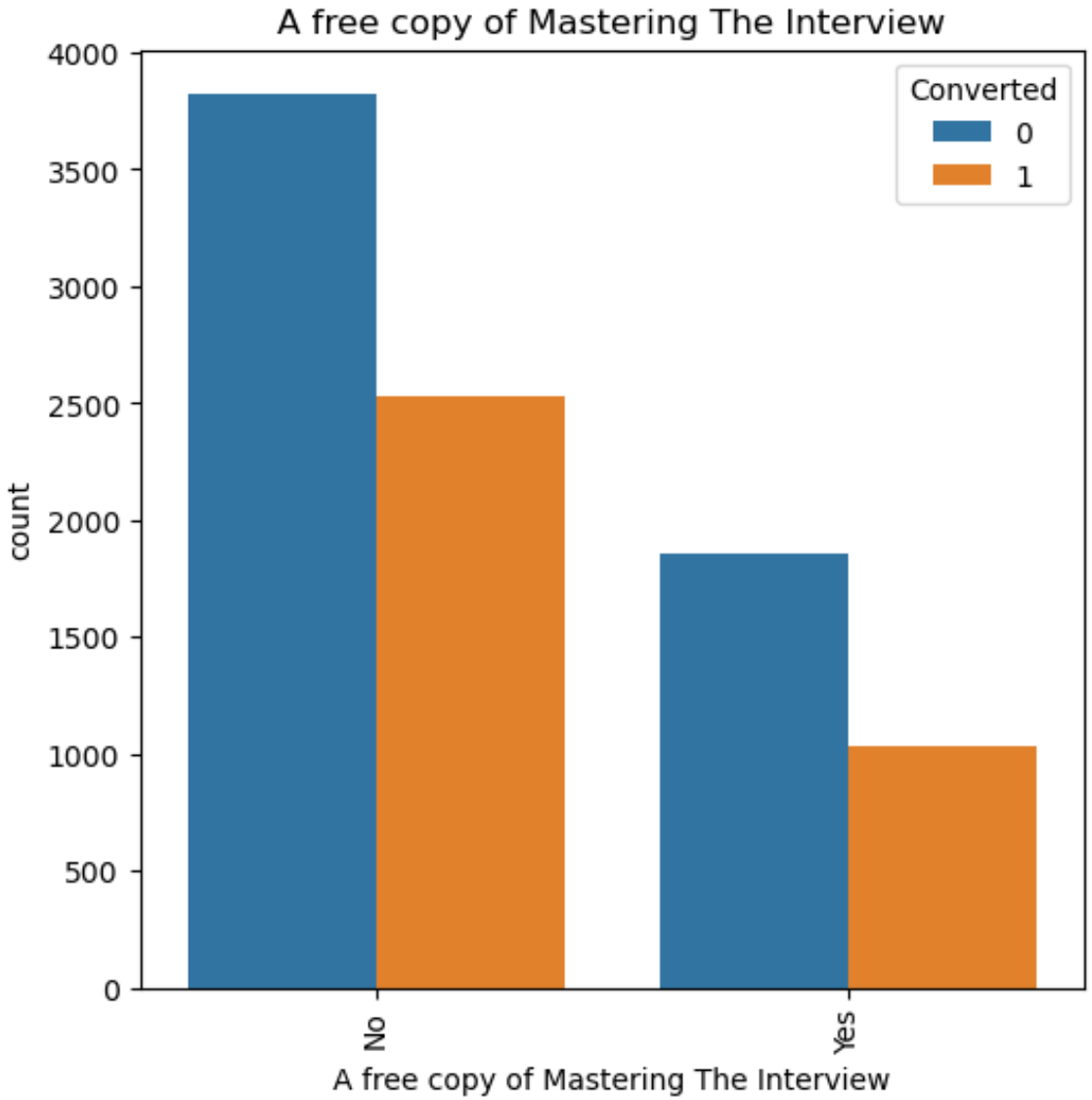
Graph comparing leads from various places or city which converted(orange) and didn't convert(blue) into potential customer



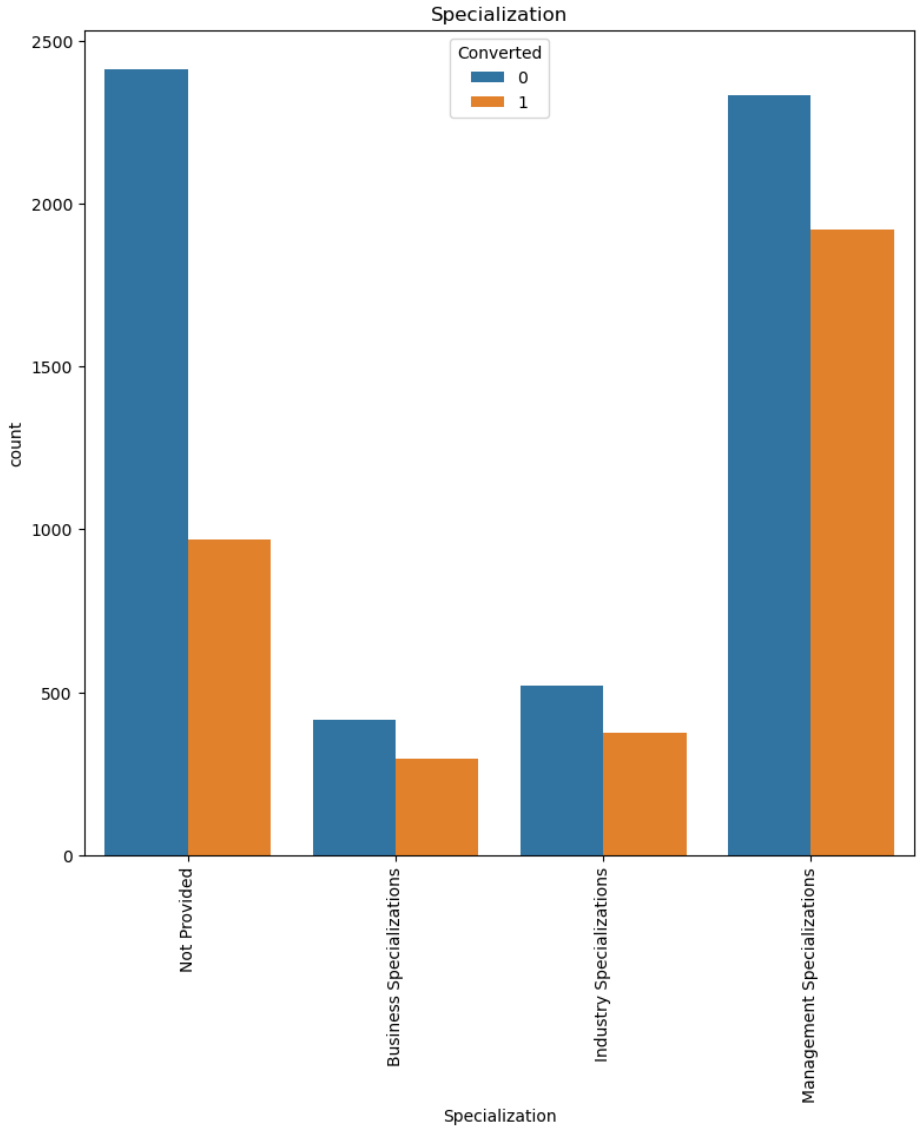
Graph comparing proportion of leads from 'Do not email' column which converted(orange) and didn't convert(blue) into potential customer



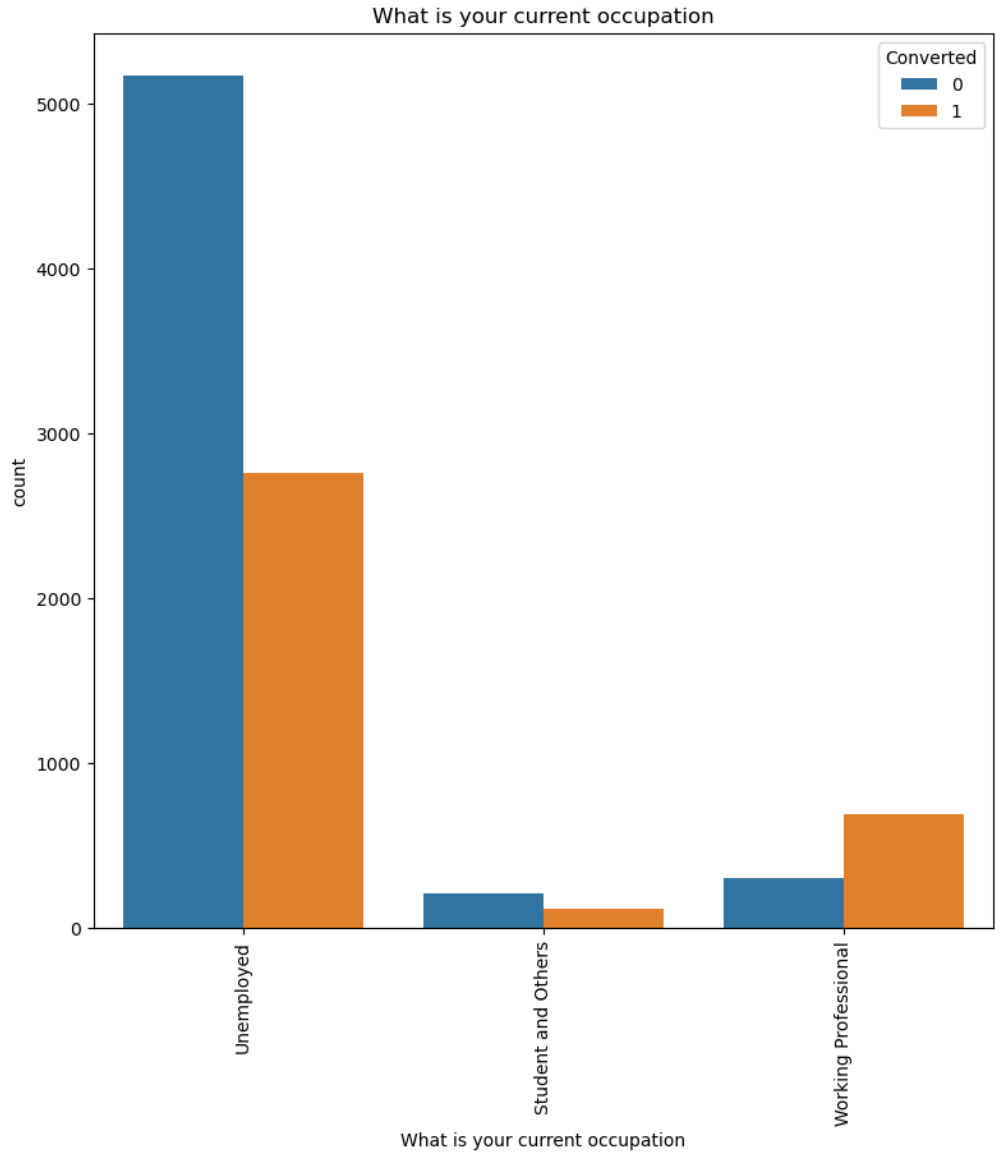
Graph comparing proportion of leads from 'A free copy of mastering the Interview' column which converted(orange) and didn't convert(blue) into potential customer



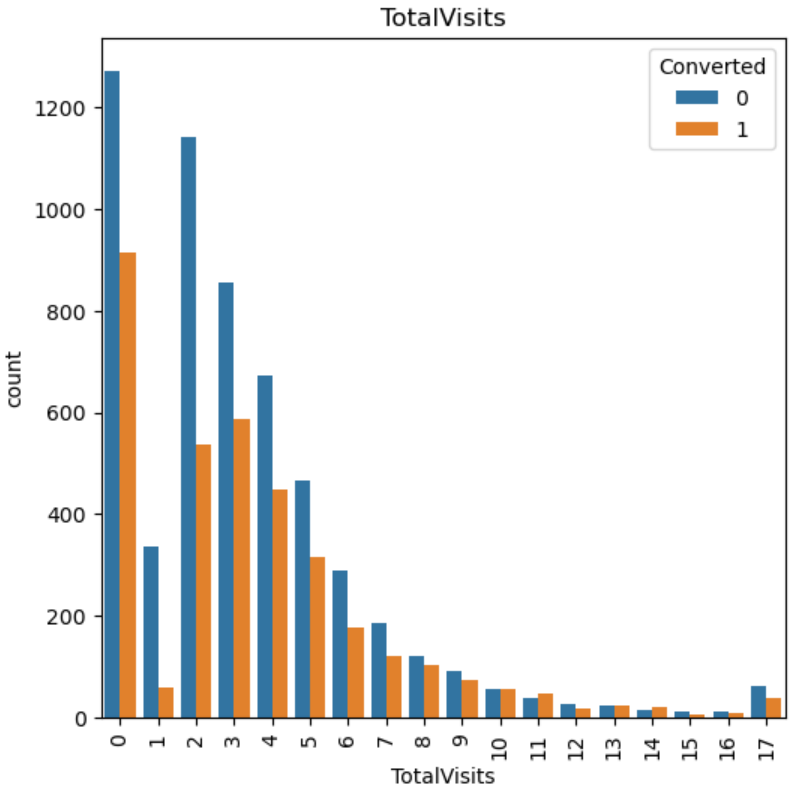
Graph comparing proportion of leads from ‘Specialization’ column which converted(orange) and didn’t convert(blue) into potential customer

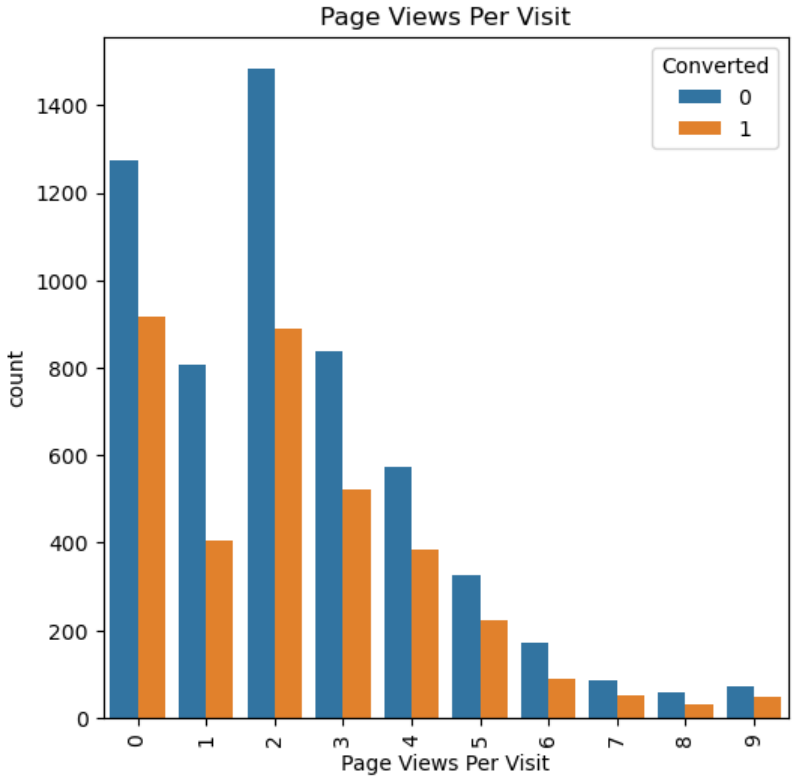


Graph comparing proportion of leads from 'What is your current occupation' column which converted(orange) and didn't convert(blue) into potential customer

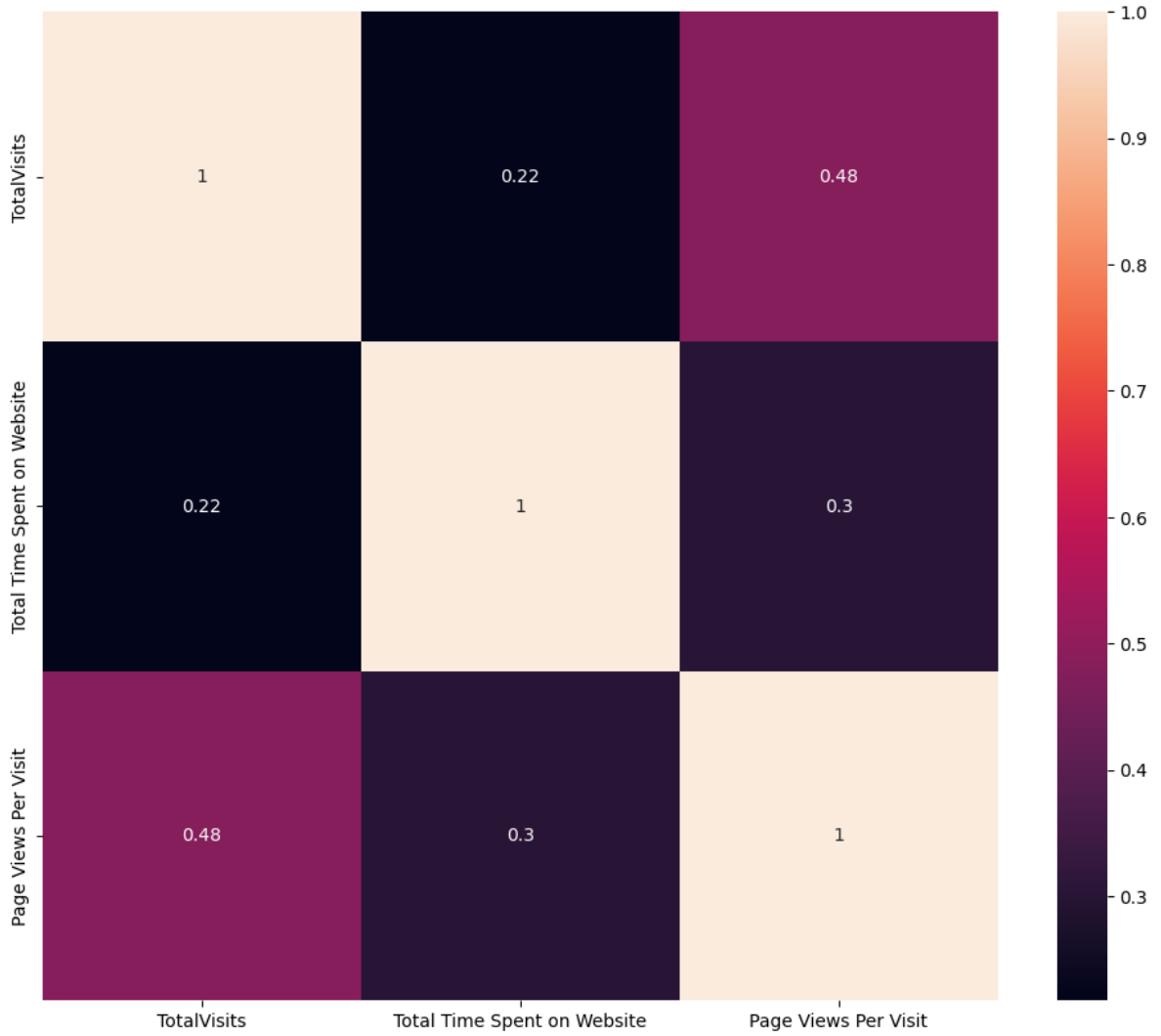


Graph comparing proportion of leads from numerical columns which converted(orange) and didn't convert(blue) into potential customer

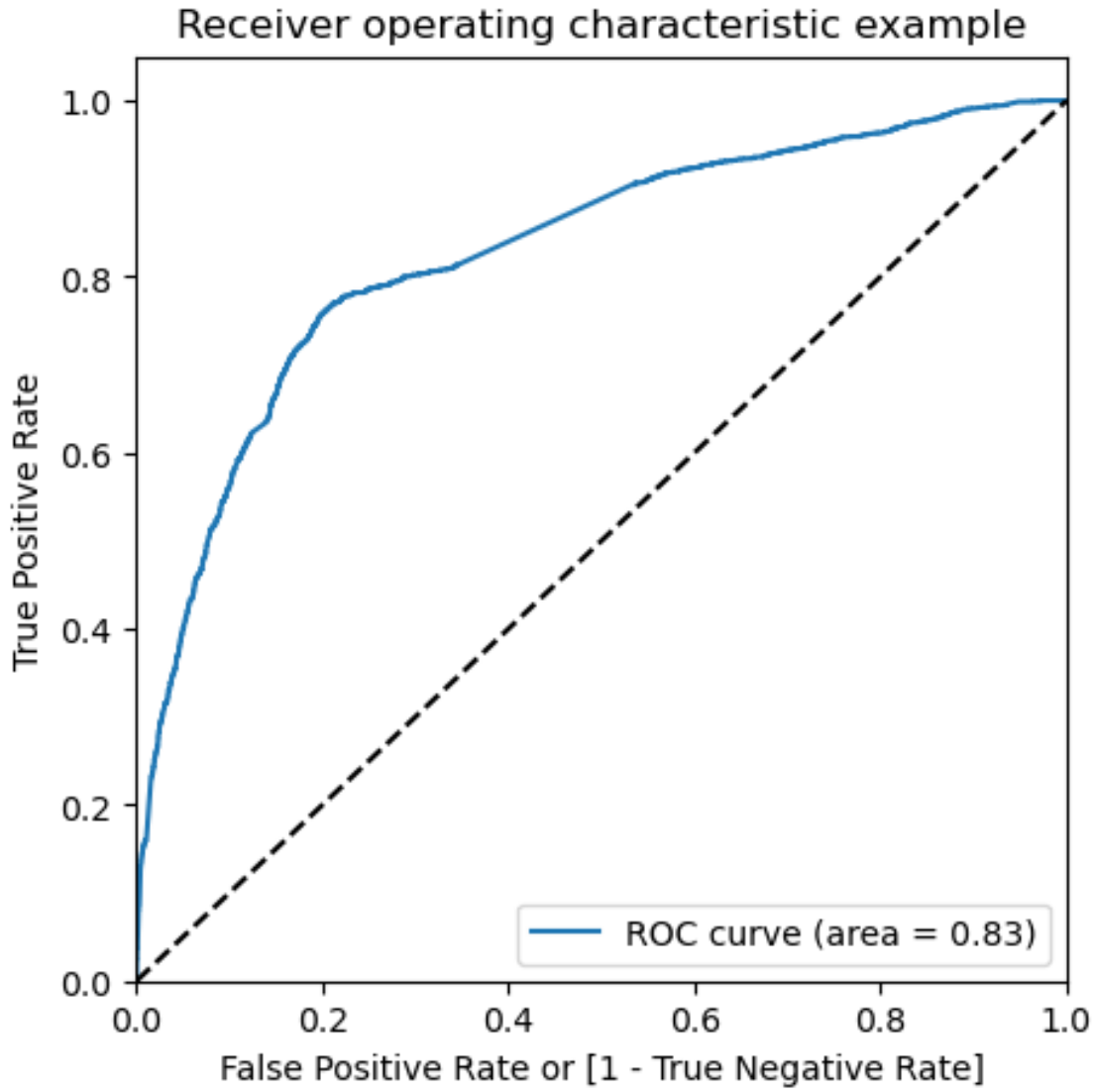




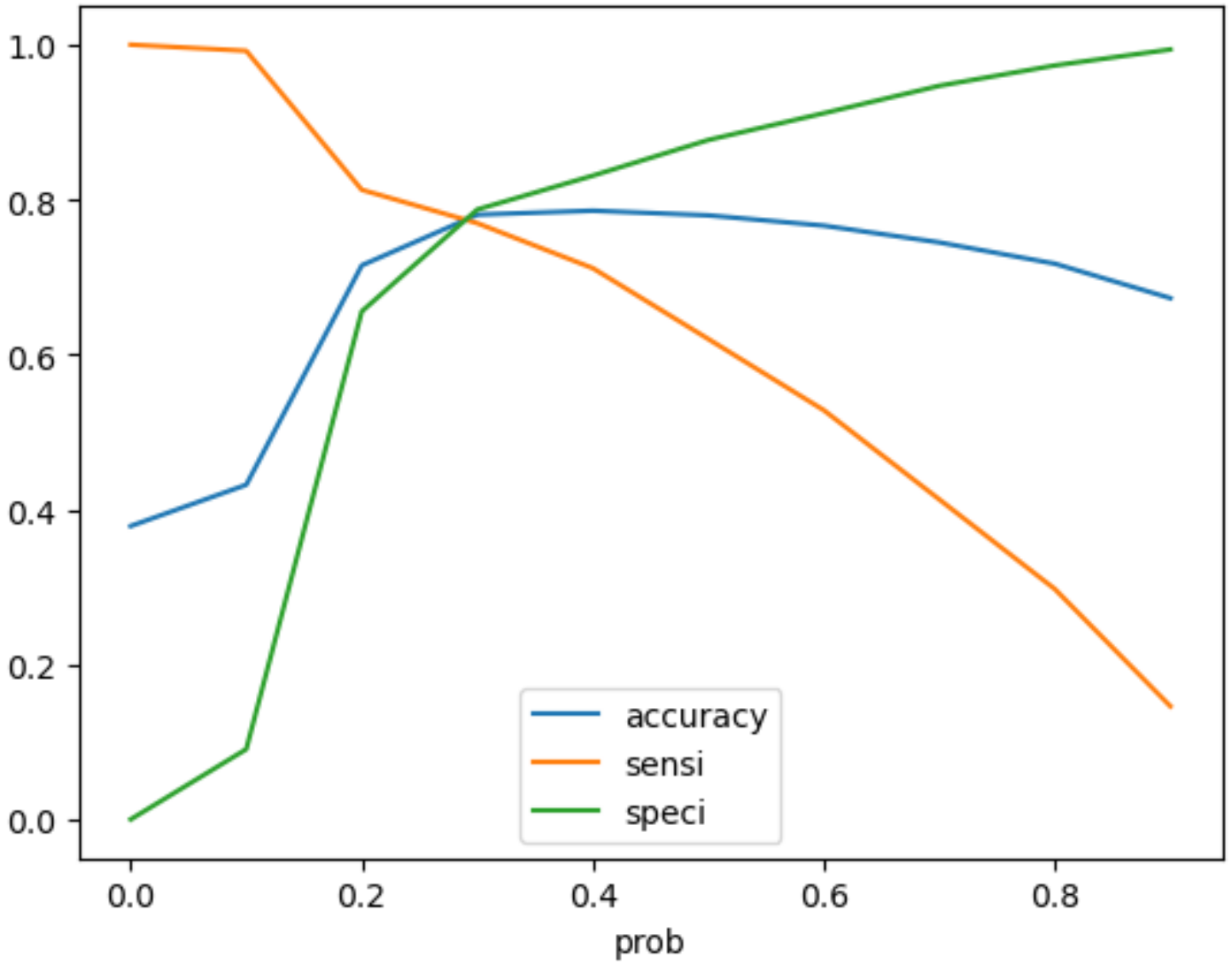
Heatmap checking collinearity in numerical columns



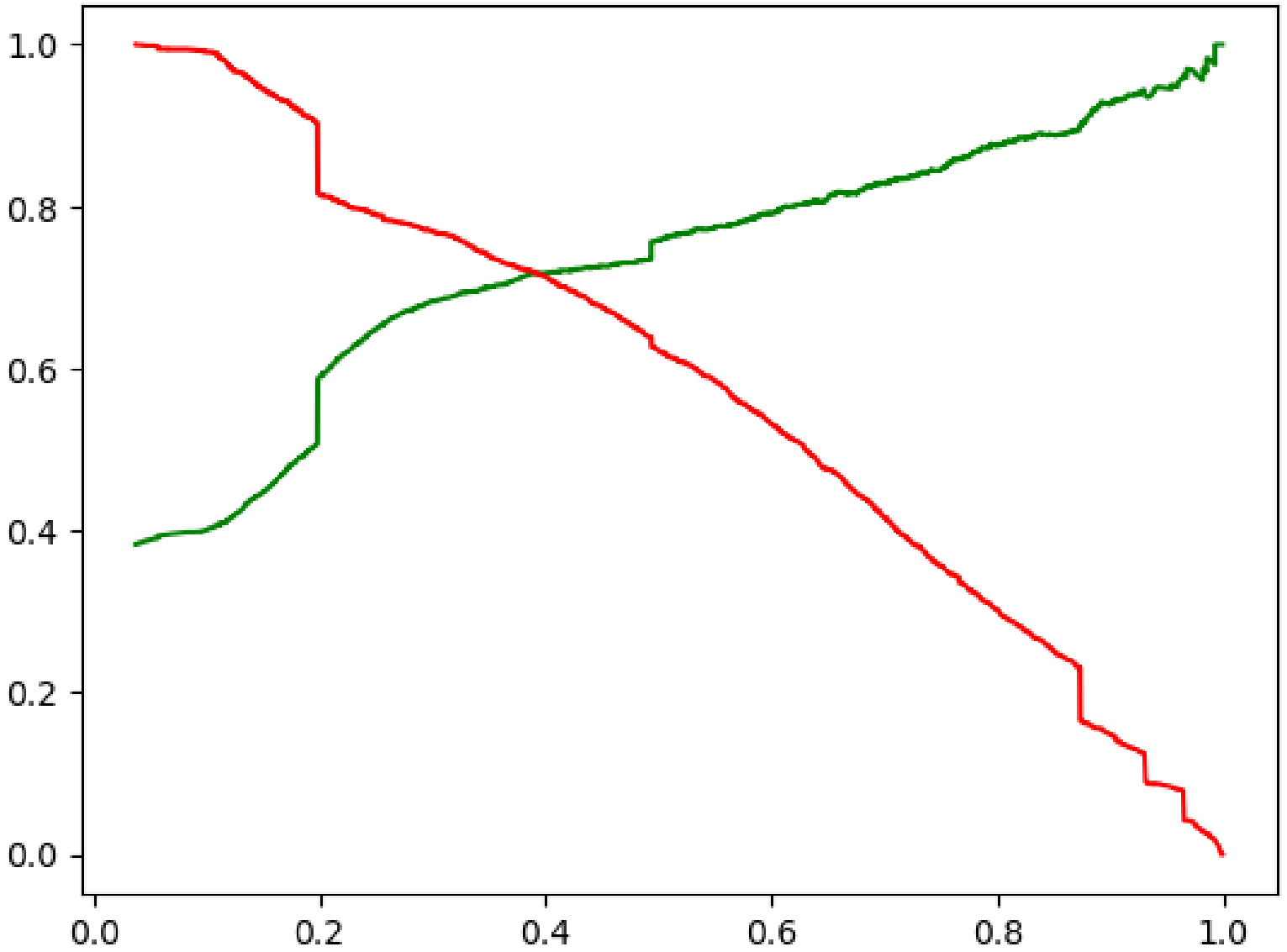
The area under ROC curve is 0.83, which is quite good



We got cutoff of 0.3 for sensitivity and specificity



While for precision-recall tradeoff we got cutoff value of 0.304



ANALYSIS OF MODEL

For training set at cut off of 0.3 we got

- Accuracy : 78.24%
- Sensitivity : 77.26%
- Specificity : 78.83%

And for test set

- Accuracy : 77.86%
- Sensitivity : 75.21%
- Specificity : 79.61%

For training set, with the precision-recall tradeoff of 0.304, we got

- Accuracy :78.37%
- Precision : 69.23%
- Recall : 77.18%

And for test set:

- Accuracy :77.82%
- Precision : 70.95%
- Recall : 74.98%

CONCLUSION

- We got precision-recall tradeoff at 0.304, which gave us accuracy of 77.82%, Precision of 70.95% and Recall of 74.98% on test sets
- By increasing the recall, X Education can identify a larger proportion of potential leads who are likely to convert, and focus their phone calls and efforts on these individuals. This can lead to a higher conversion rate and more efficient use of resources
- To increase recall, X Education can consider adjusting the model's decision threshold. By lowering the threshold, the model will classify more cases as positive, potentially capturing more of the potential leads

VARIABLES THAT MATTERED MOST: (in descending order)

- Total Time Spent on Website
- Lead Origin_Lead Add Form
- Lead Origin_Lead Import
- Page Views Per Visit
- Do Not Email
- Lead Source_Olark Chat
- What is your current occupation_Working Professional
- Lead Origin_Landing Page Submission
- Lead Source_Reference
- Specialization_Industry Specializations
- A free copy of Mastering The Interview
- TotalVisits

SUGGESTIONS

1) **Total Time Spent on Website** has a positive coefficient, indicating that the more time a customer spends on the website, the higher the probability of conversion. Therefore, the company should focus on improving the user experience of their website to make it more engaging and informative.
2) **Lead Origin_Lead Add Form** has a positive coefficient, indicating that customers who come through the Lead Add Form have a higher probability of conversion. The company should try to create more opportunities for potential customers to fill out the Lead Add Form on their website.
3) **Lead Origin_Lead Import** has a negative coefficient, indicating that leads imported from external sources have a lower probability of conversion. The company should focus more on generating their own leads through marketing campaigns and other means.
4) **Page Views Per Visit** has a negative coefficient, indicating that customers who visit fewer pages on the website have a higher probability of conversion. This suggests that the company should focus on making their website more user-friendly and easier to navigate.
5) **Do Not Email** has a negative coefficient, indicating that customers who opt-out of emails have a higher probability of conversion. Therefore, the company should respect the privacy choices of their customers and avoid sending unnecessary emails.
6) **Lead Source_Olark Chat** has a negative coefficient, indicating that customers who come through Olark Chat have a lower probability of conversion. The company should investigate this further and find ways to make the chat experience more engaging and helpful for potential customers.

SUGGESTIONS

6) **What is your current occupation_Working Professional** has a positive coefficient, indicating that working professionals have a higher probability of conversion. Therefore, the company should create targeted marketing campaigns aimed at this demographic.
7) **Lead Origin_Landing Page Submission** has a negative coefficient, indicating that customers who come through landing page submissions have a lower probability of conversion. The company should try to make the landing pages more informative and engaging to improve the conversion rate.
8) **Lead Source_Reference** has a negative coefficient, indicating that customers who come through reference have a lower probability of conversion. The company should investigate this further and find ways to improve the referral process.
9) **A free copy of Mastering The Interview** has a negative coefficient, indicating that customers who opt-out of receiving the free book have a higher probability of conversion. Therefore, the company should avoid making this a mandatory option and respect the choices of their customers.
10) **Lead Source_Organic Search** has a negative coefficient, indicating that customers who come through organic search have a lower probability of conversion. The company should investigate this further and find ways to improve their SEO strategy to attract more high-quality traffic.
11) **TotalVisits** has a negative coefficient, indicating that customers who visit fewer pages on the website have a higher probability of conversion. This suggests that the company should focus on making their website more informative and engaging to keep the customers interested and exploring more pages.

THANKYOU