

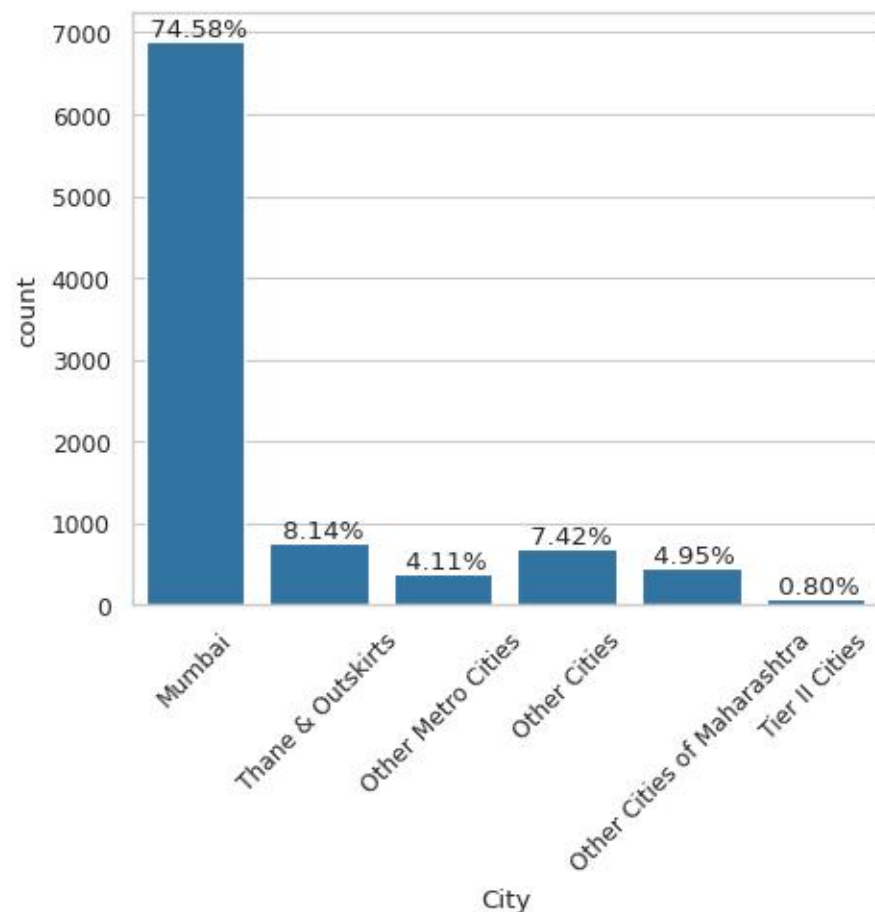
LEAD SCORING CASE STUDY

SUBMITTED BY-

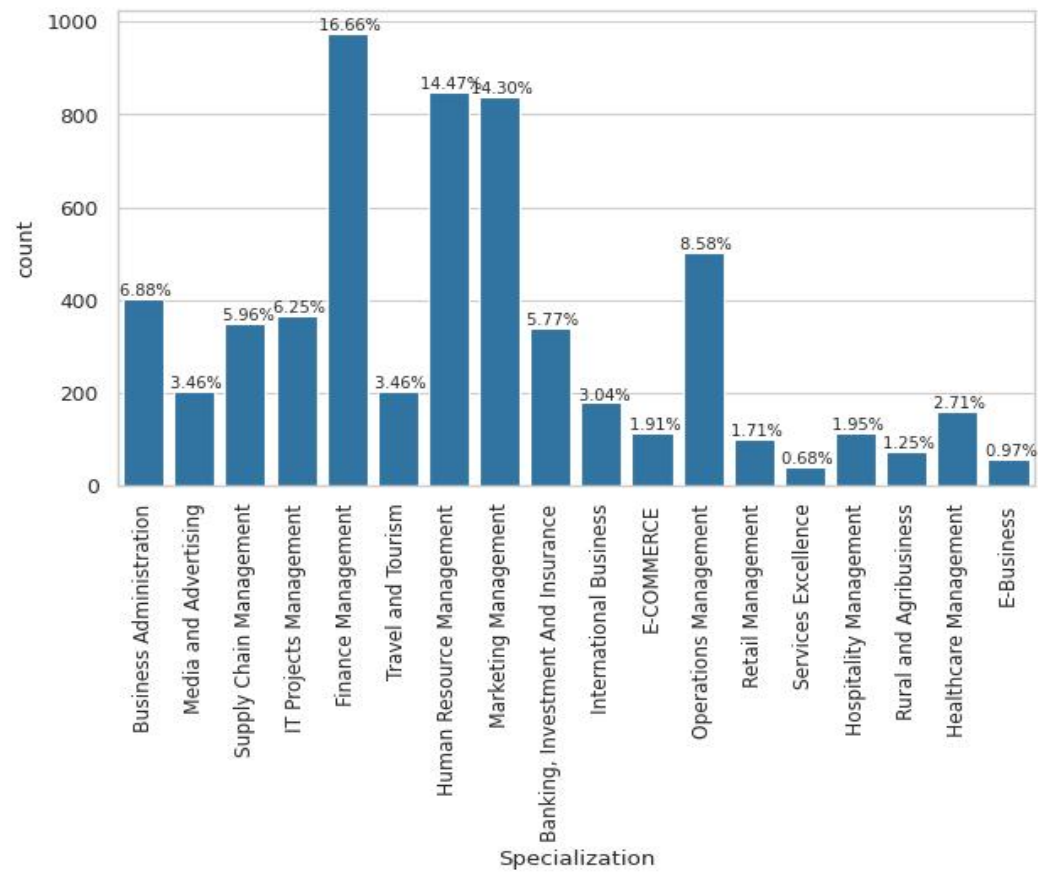
K DHEERAJ KUMAR

EMAIL: dhrrj511@gmail.com

Data cleaning and EDA

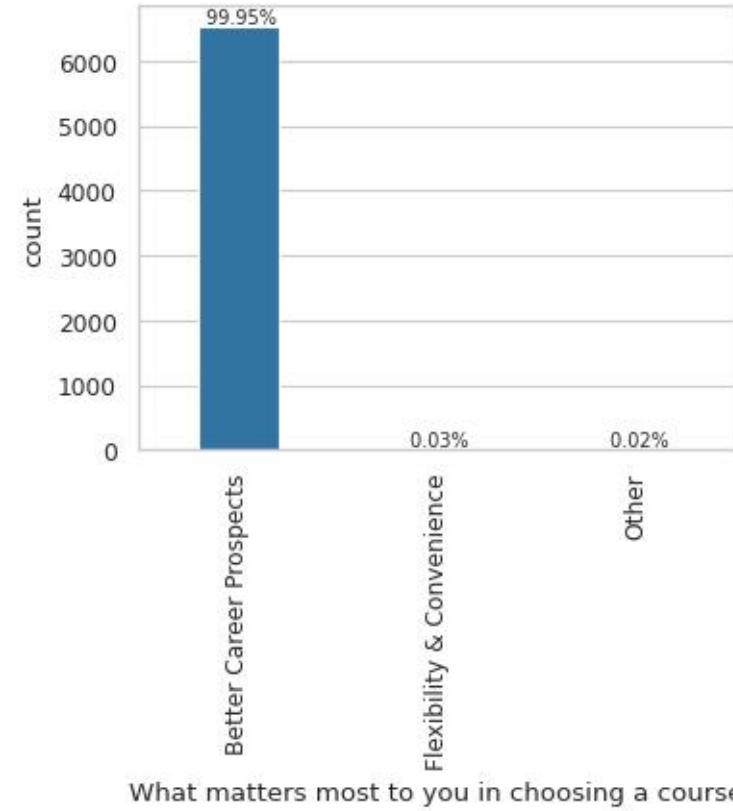
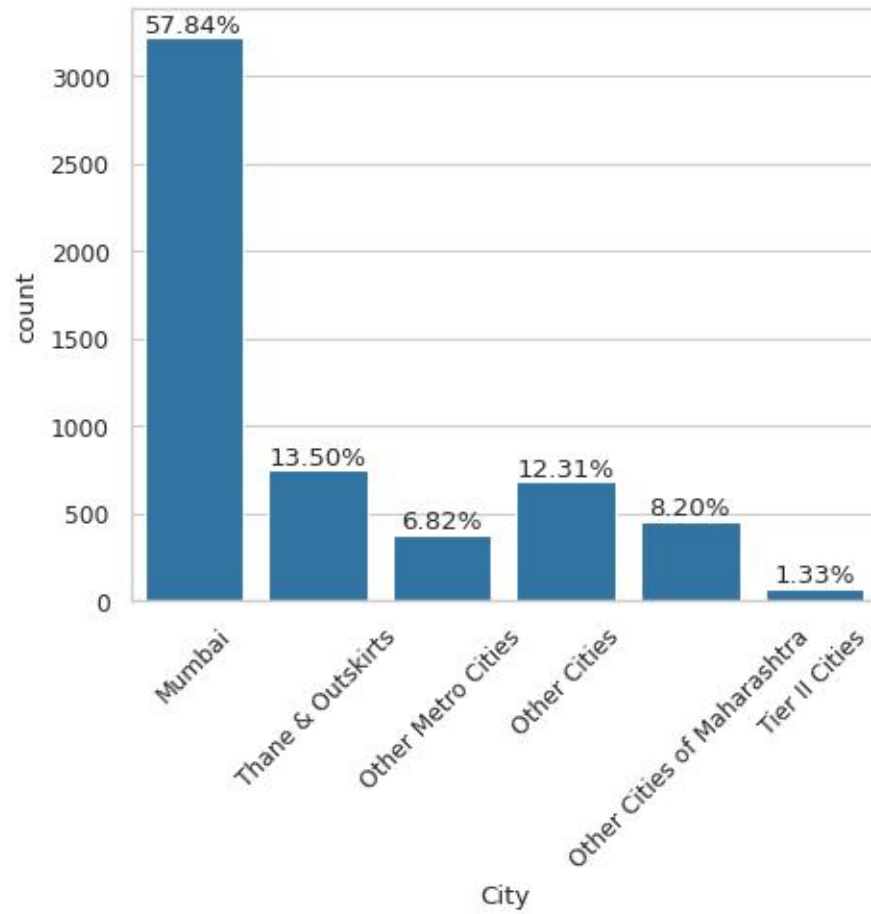


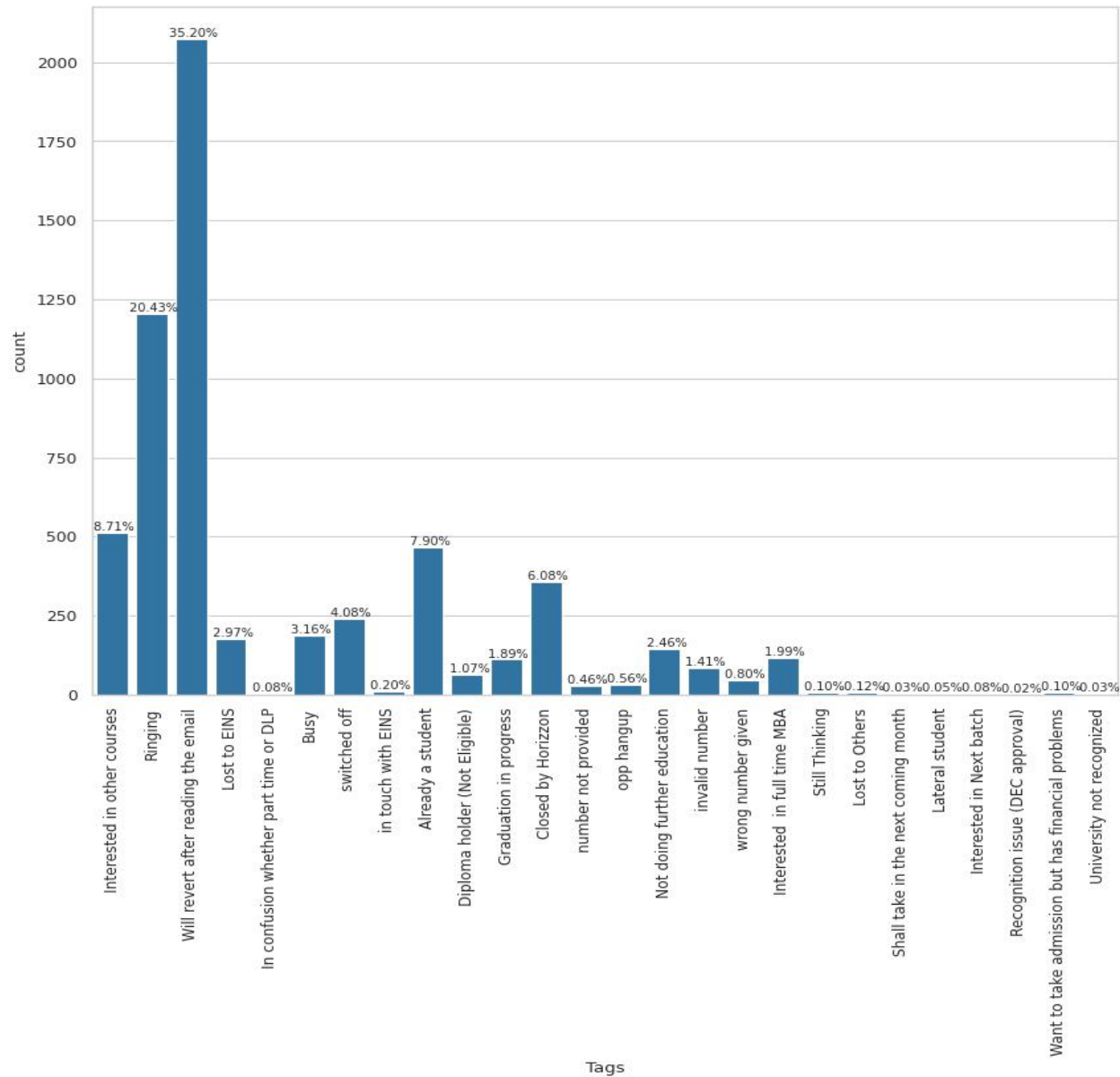
- Note: % values are calculated from total number of rows of the df
- Most values belong to mumbai
- so, we can impute missing values with mumba



- There are many specializations present here and the missing values are also quite high, therefore we can create another category 'others'

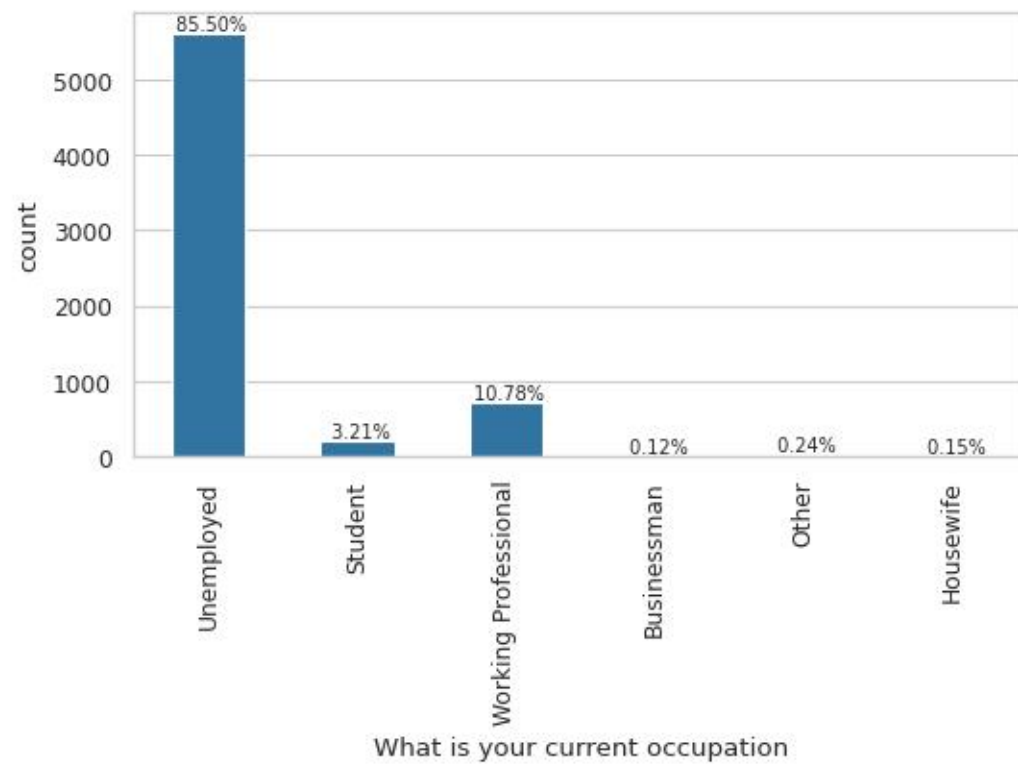
data cleaning



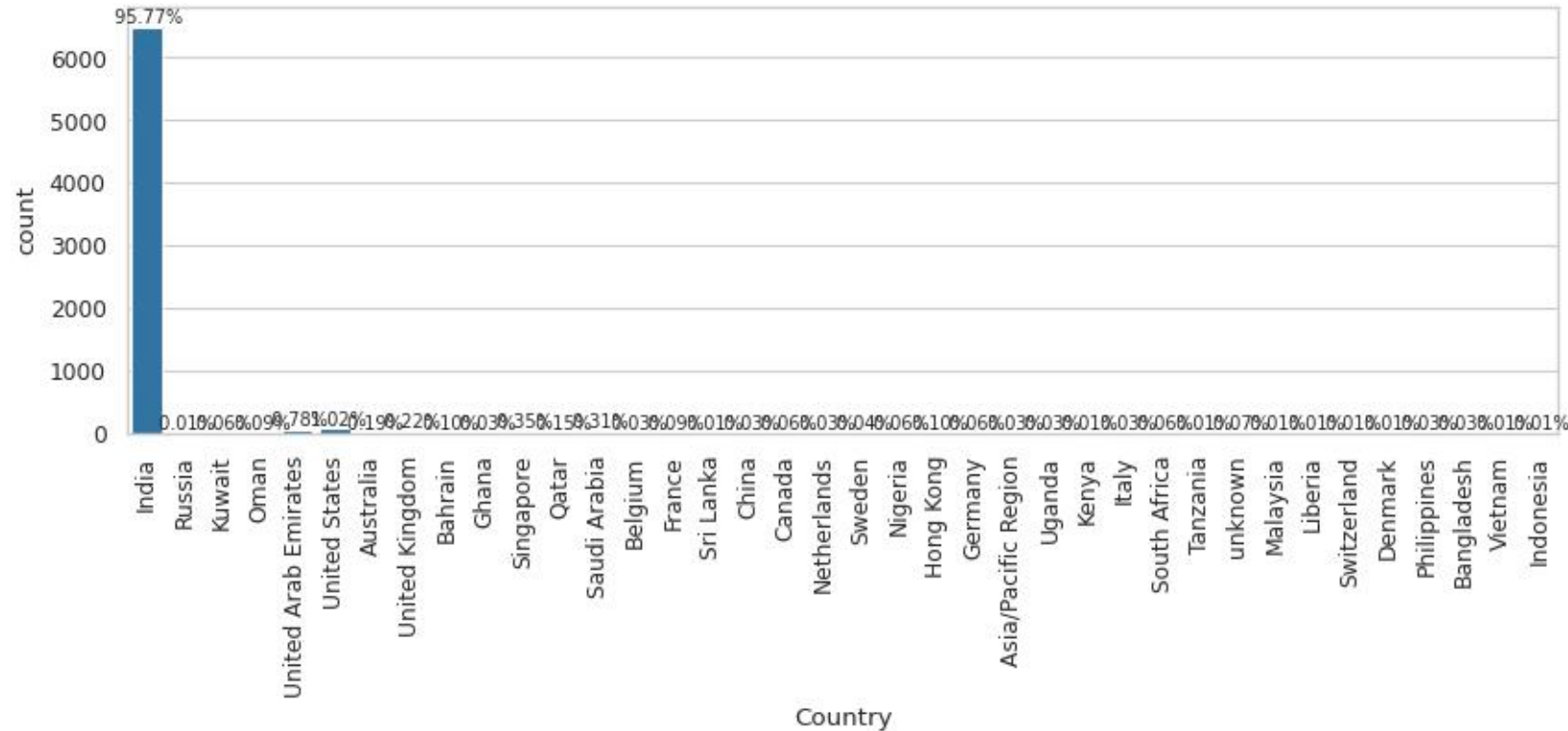


most of the values are 'we will revert after reading the email'

so we can impute the missing values with this

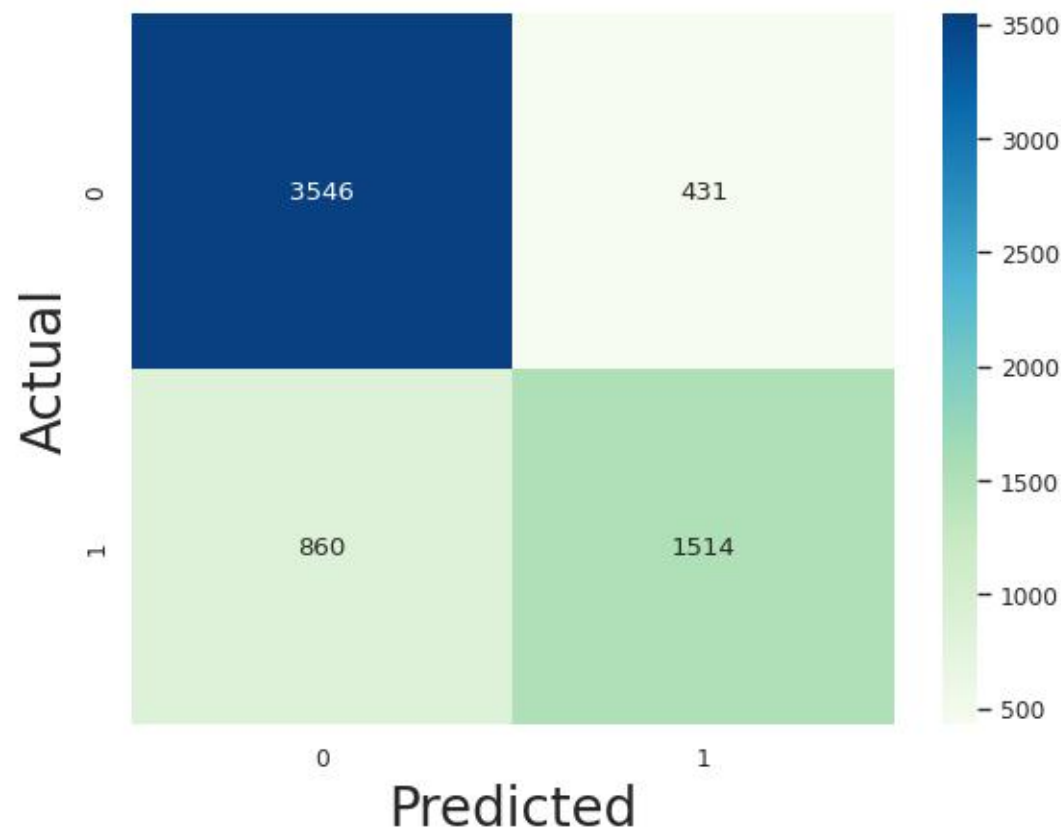


- most of them are unemployed
- so imputing with unemployed



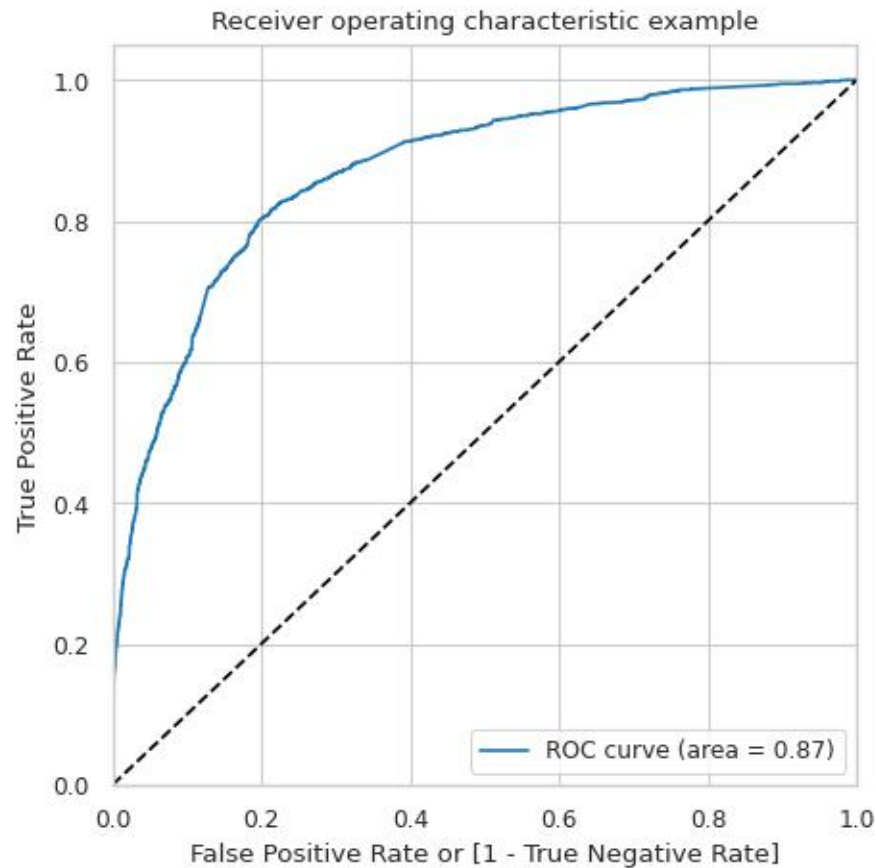
- very skewed plot, but as it is country, it could be an important variable
- so imputing the missing values with india

Confusion Matrix



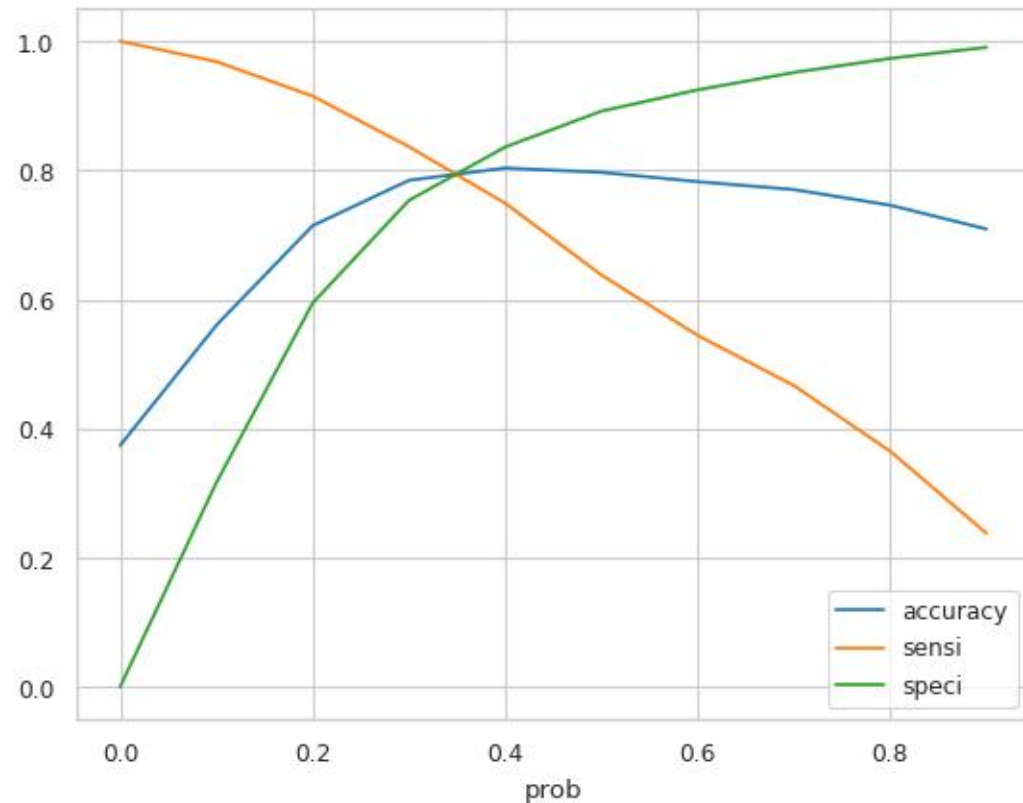
- the model performs well because true positive and true negative values are much greater than false positive and false negative values
- our model strong accuracy but struggles with recall because it misses many actual positive cases

ROC Curve



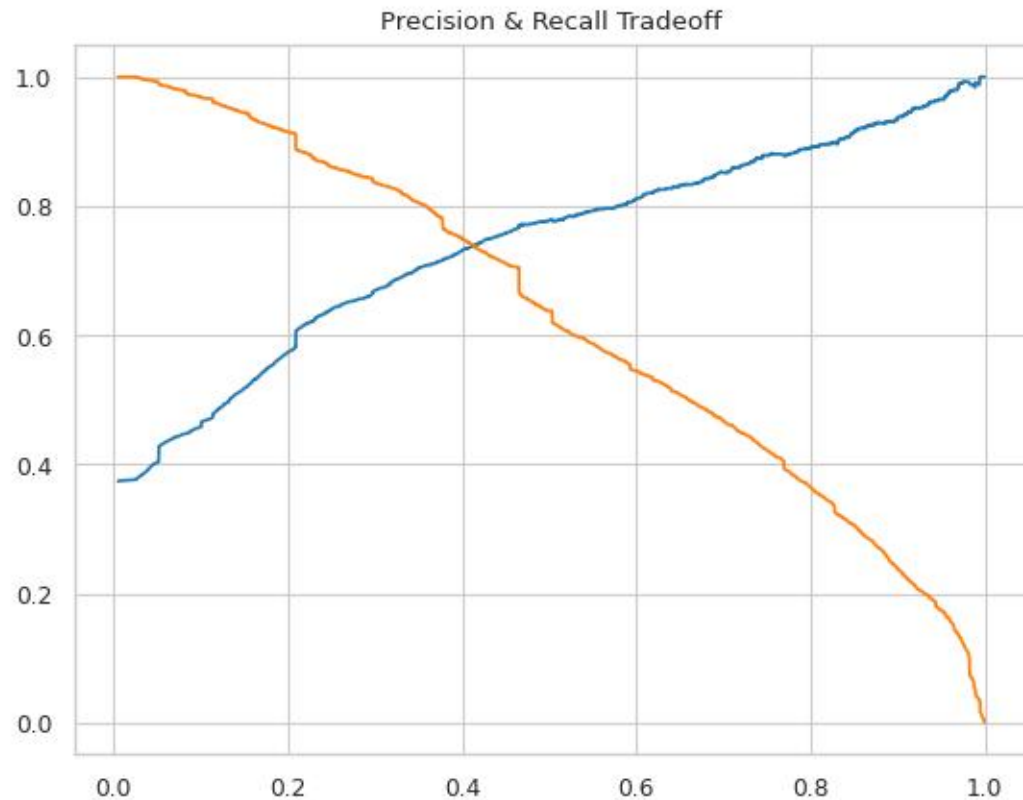
- this shows that the model is able to achieve high true positive rates with relatively low false positive rates
- the area under the curve=0.87 represents that the model has good ability to distinguish between the positive and the negative classes

Finding optimum cutoff point



- the optimal cutoff point can be seen at around 0.35
- this can be the optimal point for balancing both metrics for optimal model performance

Precision and Recall tradeoff



- Precision and recall have an inverse relationship.
- At lower thresholds, the model captures more true positives (higher recall) but also makes more false positive errors (lower precision).
- At higher thresholds, precision improves at the cost of recall

variables and their coefficients of the final model

Lead Source_Welingak Website	5.717363
What is your current occupation_Working Professional	2.795847
Last Activity_SMS Sent	2.050196
Last Activity_Others	1.283260
Last Activity_Email Link Clicked	1.239169
Total Time Spent on Website	0.962912
Last Activity_Email Opened	0.858354
What is your current occupation_Student	0.562399
Lead Source_Olark Chat	0.361796
const	0.014192
Last Notable Activity_Modified	-0.721922
Last Notable Activity_Email Link Clicked	-0.928944
Specialization_Others	-1.706168
Lead Origin_Landing Page Submission	-2.081977

Recommendations

Call these leads:

- Those from "Welingak Websites" and "Reference" because they are likely to convert.
 - Working professionals, as they are also likely to convert.
 - Leads who spent more time on the website, since they are more likely to convert.
 - Those from "Olark Chat" because they have a higher chance of converting.
 - Leads whose last activity was an SMS sent, as they are more likely to convert.
-
- Do not call these leads:
 - Those whose last activity was an "Olark Chat Conversation," as they are unlikely to convert.
 - Leads from "Landing Page Submission," since they are also unlikely to convert.
 - Leads whose specialization is "Others," as they are less likely to convert.
 - Leads who chose "Do not Email" as "yes," since they are unlikely to convert.