LEAD SCORE CASE STUDY

Problem statement

- X Education sells online courses to industry professionals
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process more efficient, the company wishes to identify the most potential leads, also known as ‘Hot Leads’.
- If they successfully identify this set of 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

Objective

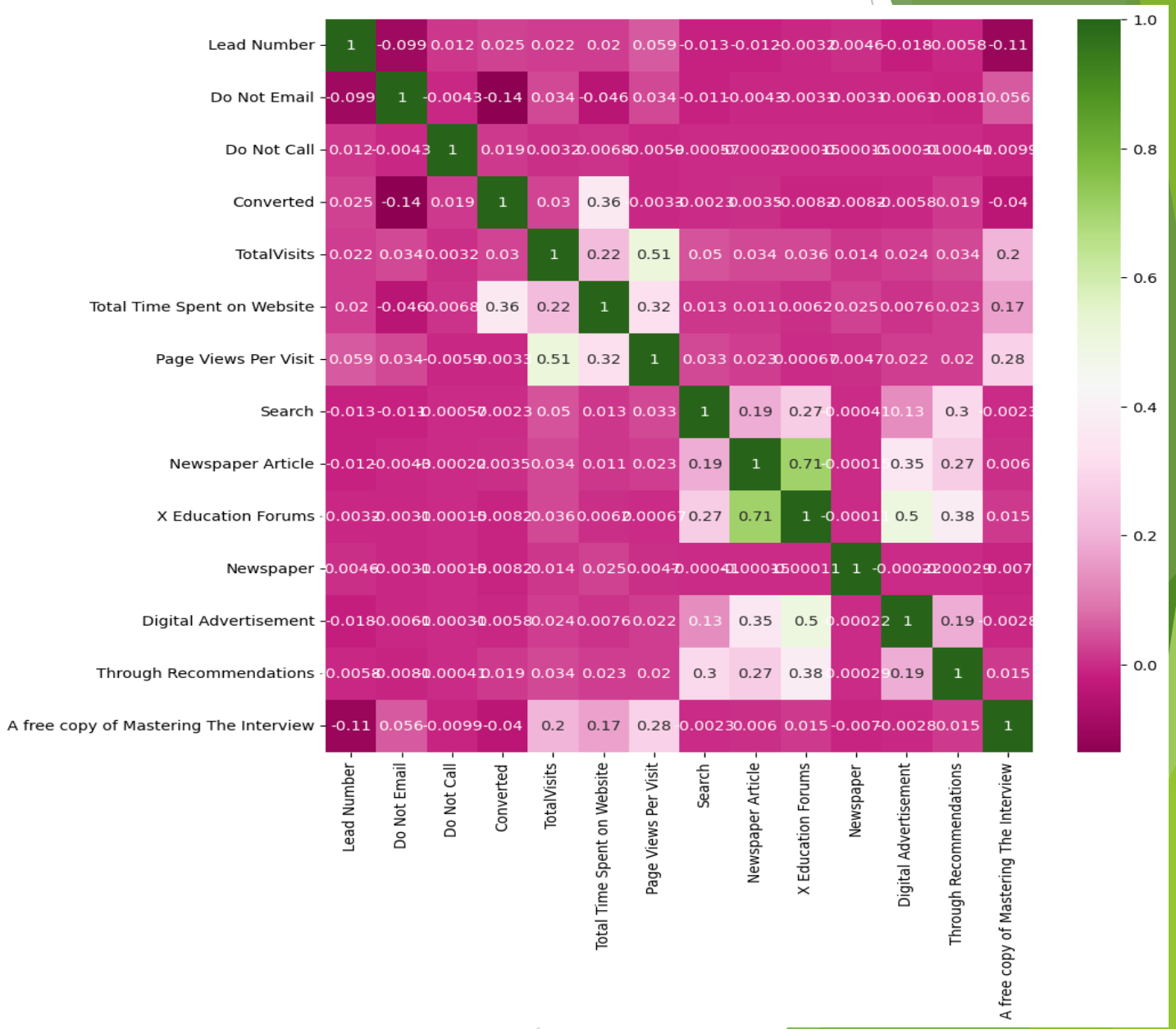
- **An education company named X Education sells online courses to industry professionals.**
- **On any given day, many professionals who are interested in the courses land on their website and browse for courses.**
- **Here we have to analyse the previous user data to find out which parameters influence the conversion rate of the course**

Solution

- Data cleaning and data manipulation
- EDA – Univariate data analysis, Bivariate data analysis
- Feature scaling , Dummy variables and encoding of data
- Logistic regression for model making and prediction
- Model presentation
- Conclusion

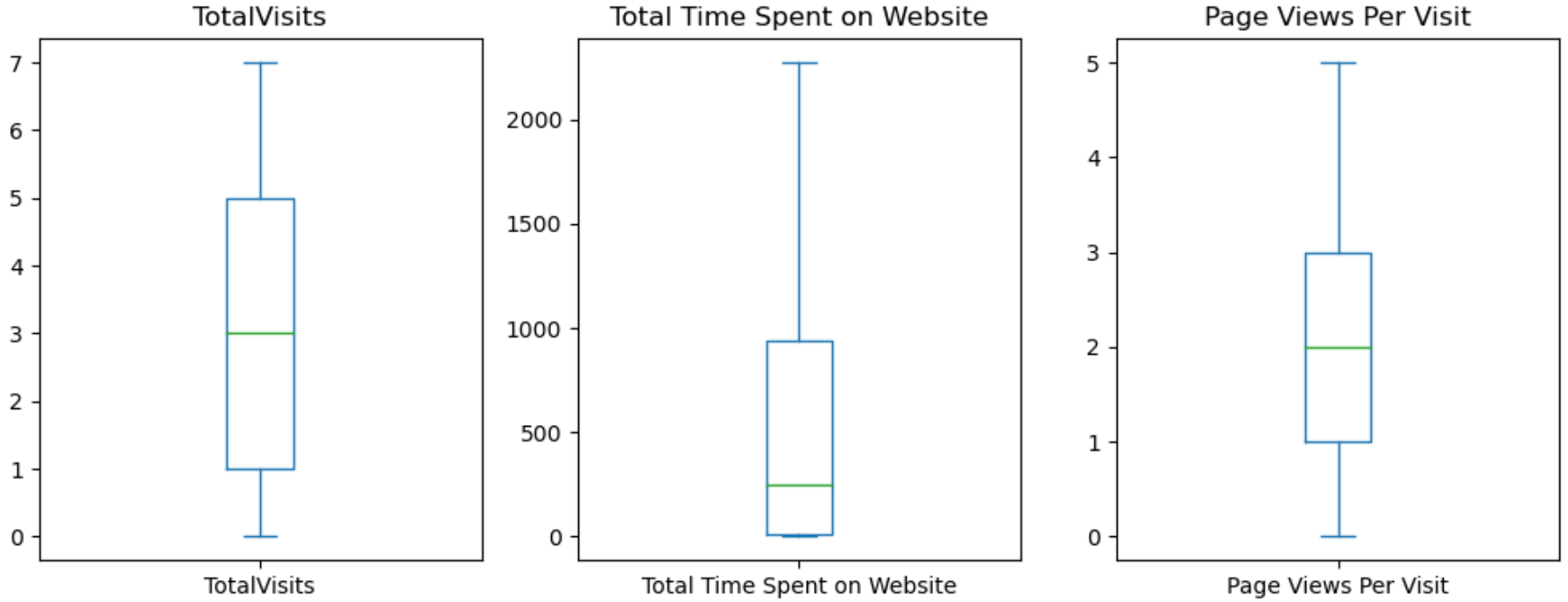
EDA

- Corelation between variables

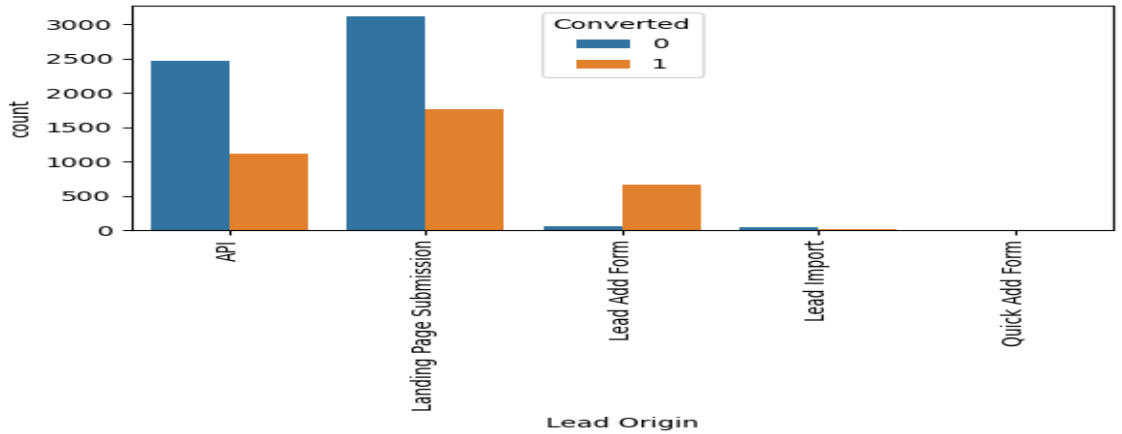


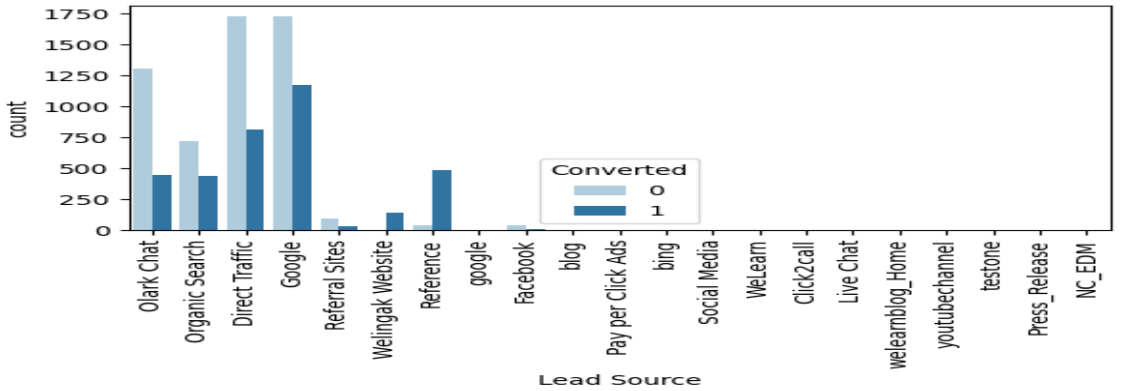
•Univariate Analysis

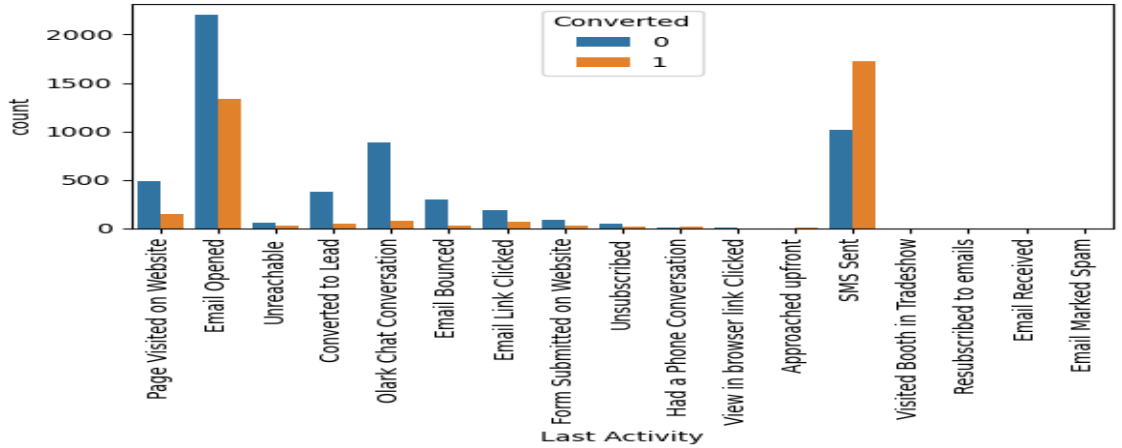
All variables seems to be truncated well.
The 50% of the Total visits of users come in the range of 1-5.
The 50% of the time spent on the website is between 0- 1000.
The 50% of the page views per visit is between 1-3.

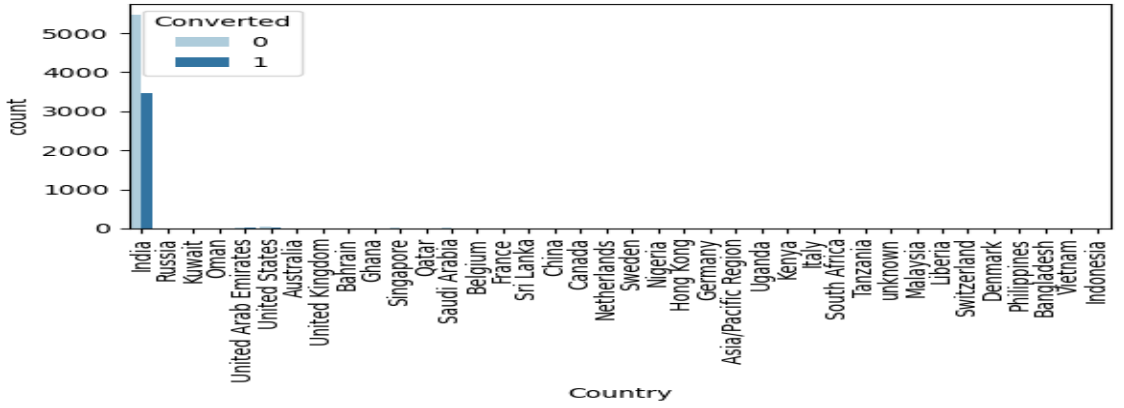


Bivariate analysis









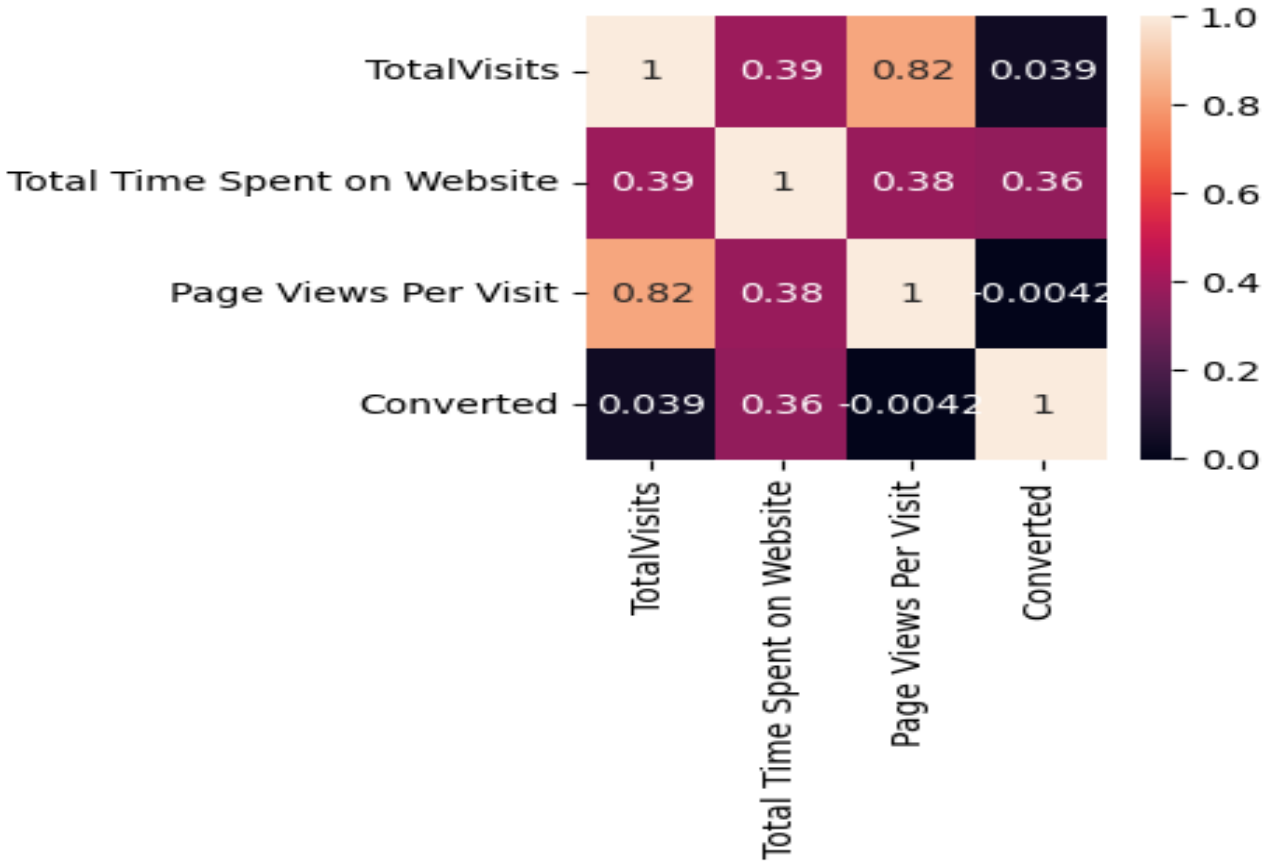
Bivariate analysis

- More conversion is observed on Landing page submission as Lead origin.
- Data traffic and Google are relatively better lead sources.
- SMS sent and Email opened as Last activity have high number of conversion compared to others.
- Almost all the converted users are from India. So There is no valuable insight from that variable.
- People specialized more in Finance management are more likely to convert.
- Unemployed users are more interested in taking the course than working professiionals.
- Almost all users select the course for better career prospects.
- Users who will revert after reading the Email are very likely to convert, but users who keeps avoiding the phone calls are very less likely to convert.
- Most converted users are from Mumbai city.
- Users who sent SMS and the one's whose Email opened as their last activity are more likely to convert.

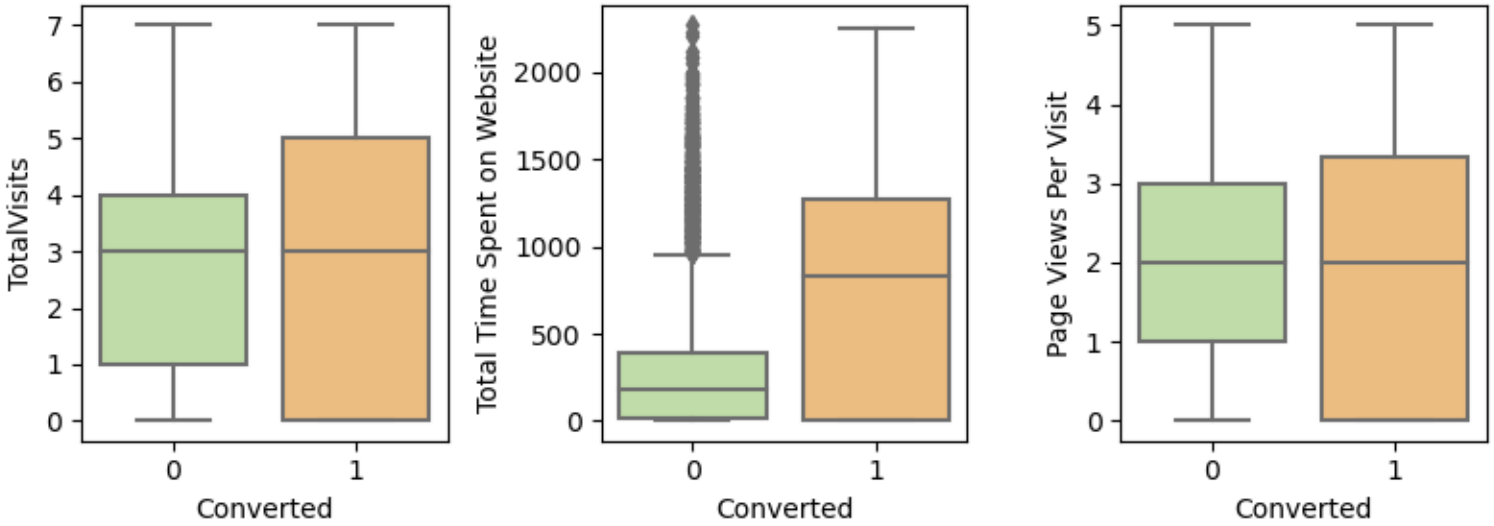
•Bivariate analysis on numerical variables

Only "Time spent on website" have a positive correlation among the numerical variables.

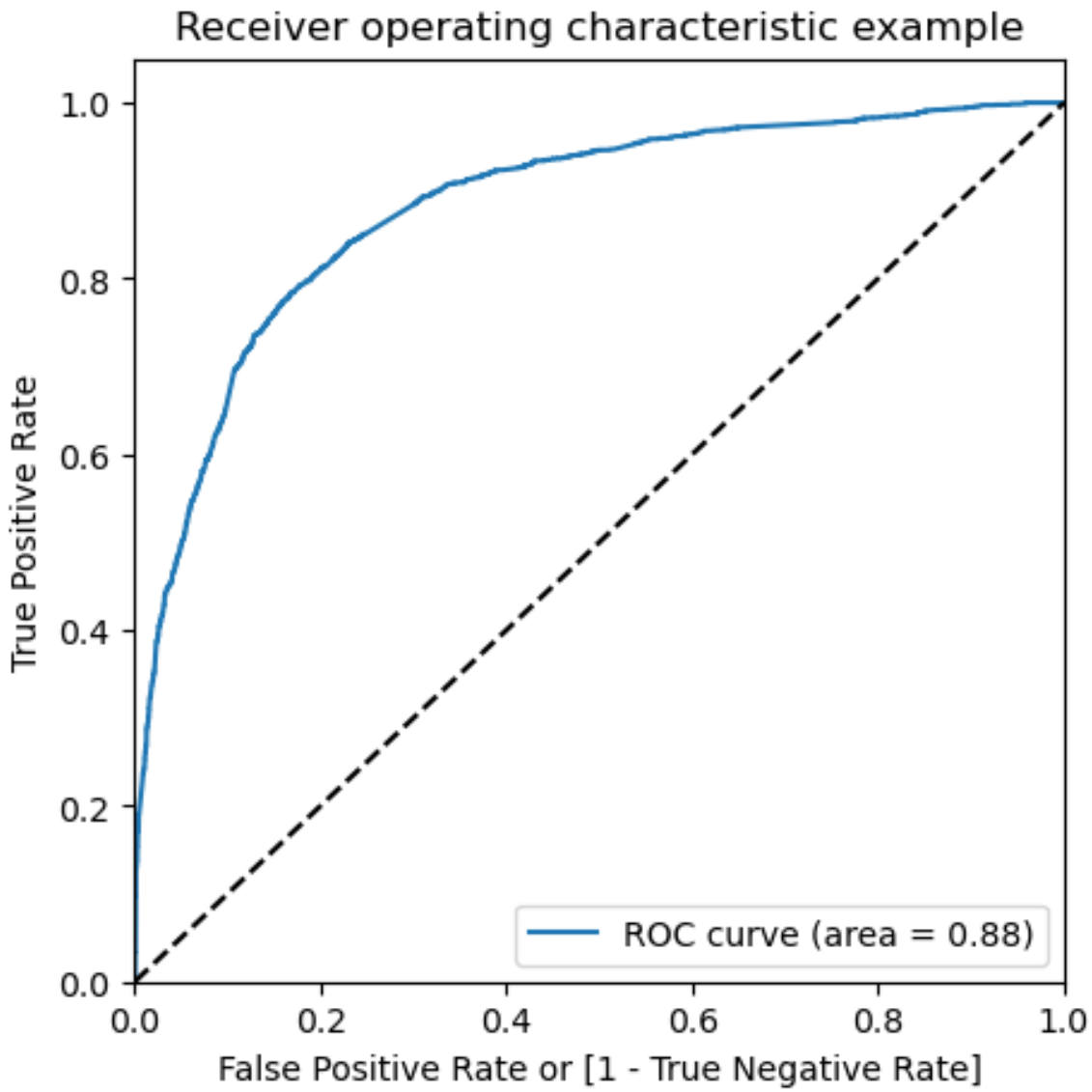
Other variableas have colinearity with each others, So they can be handled in the modelling part.



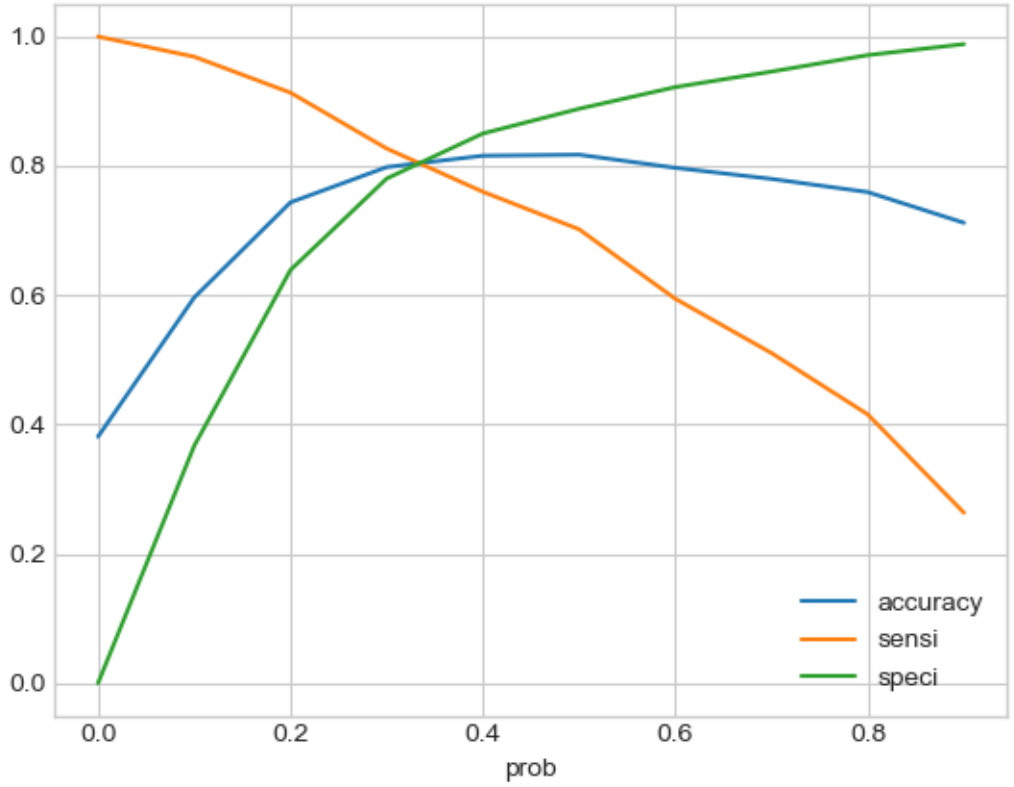
Distribution of numerical variables



ROC Curve


Receiver operating characteristic example
True Positive Rate
False Positive Rate or [1 - True Negative Rate]
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.88)

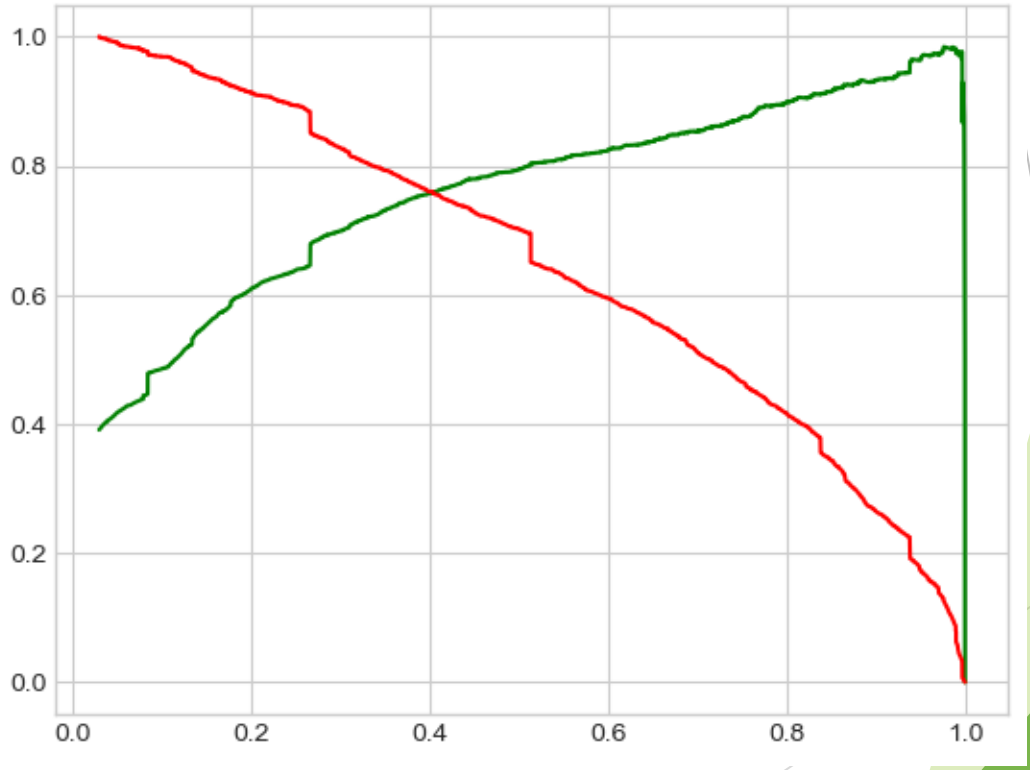
The optimal point lies in the probability 0.33. We can use this point for better representation of model



•Precision and recall tradeoff

Curve looks good.

The optimal point according to the curve is 0.4



- **Final observations from the train and test data analysis**
- **Train data**
  - Accuracy = 0.806
  - Sensitivity = 0.804
  - Specificity = 0.807
- **Test data**
  - Accuracy = 0.805
  - Sensitivity = 0.797
  - Specificity = 0.810

Conclusions and Recommendations

- Both the Train and Test data have predicted with almost comparable accuracy, specificity and sensitivity. So this model can be finalized.
- Company should focus more on leads from form fillups.
- Company should focus more on Working pofessionals and unemployed users.
- Lead source from Wellingak website should be focussed more.
- Users who spend more time on the website are more likely to convert. So keep an eye on them.
- Company should focus less on users, whose Email got bounced as the Email ID given are most probabily wrong.
- Conmpany should focus less on users with specialization of Hospital management and focus more on users of Financial management.
- Users whose last online activity is Olark chat conversation should be focussed less.