LEAD SCORING CASE STUDY

By Kumar Nilotpal,

K Rajesh

and

Keshang Yolmo

PROBLEM STATEMENT

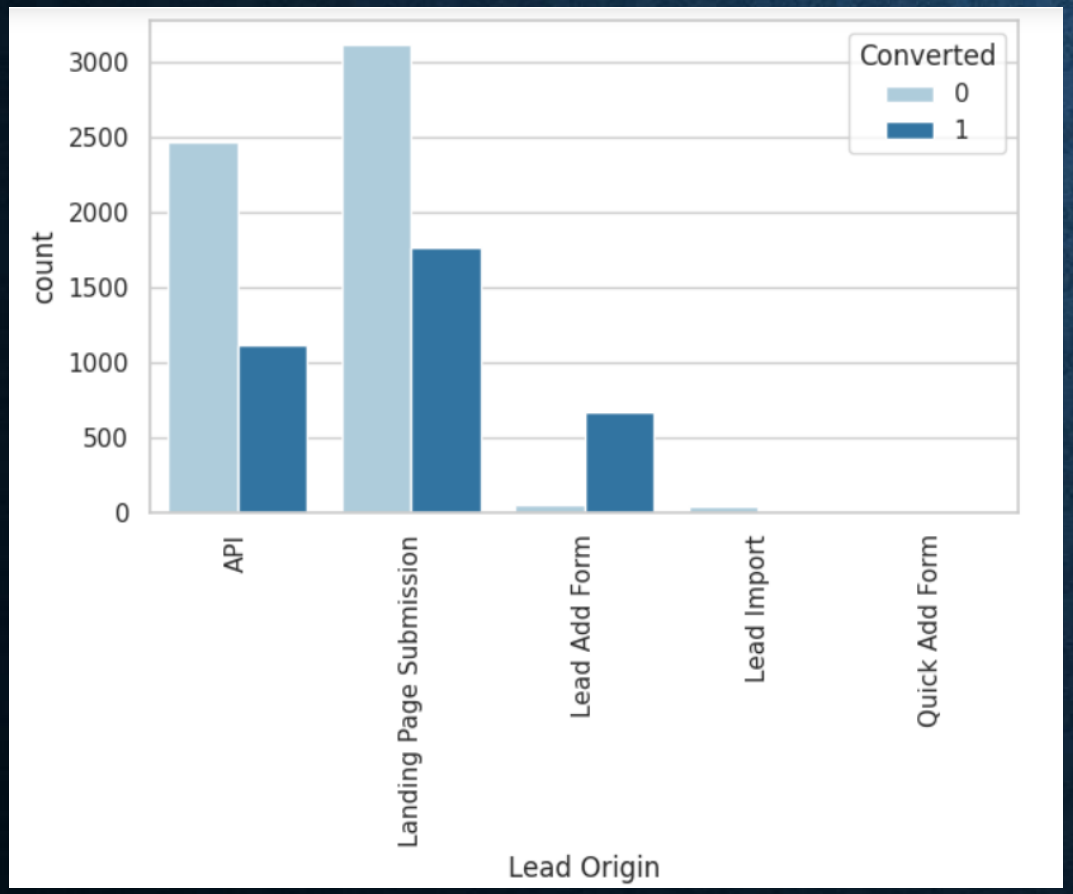
- An education company named X Education which provides online courses for industry professional. The company marks its courses on several popular websites like google.
- Company wants to select most promising leads that can be converted to paying customers.
- The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos.When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. 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.

STRATEGY

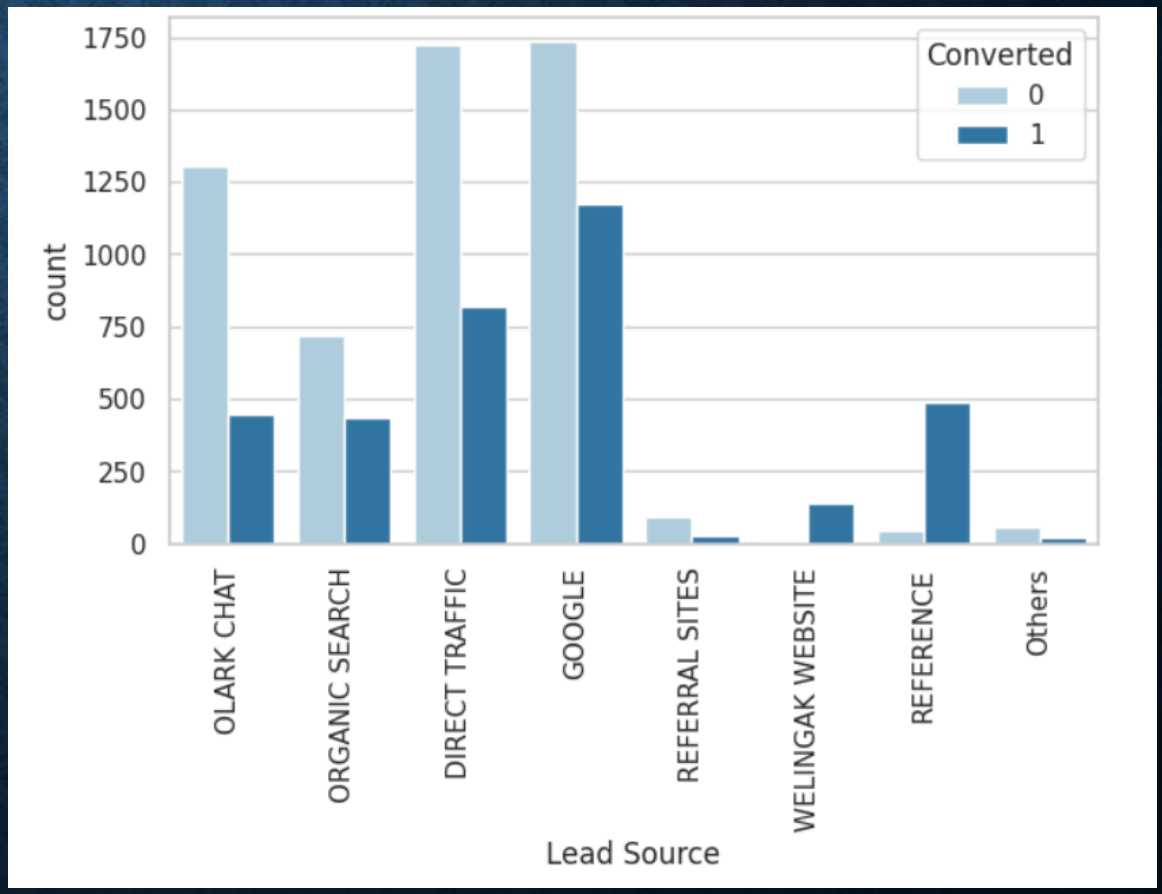
- Import required libraries
- Import Dataset
- Clean and prepare the data for further analysis
- Exploratory data analysis
- Scaling features
- Prepare the data fir model building
- Build the logistic regression model
- Assign a lead score for each leads
- Test the model on train set
- Evaluate model by different measures and metrics
- Test the model on test set
- Measure the accuracy of the model and other metrics for evaluation

DATA VISUALISATION

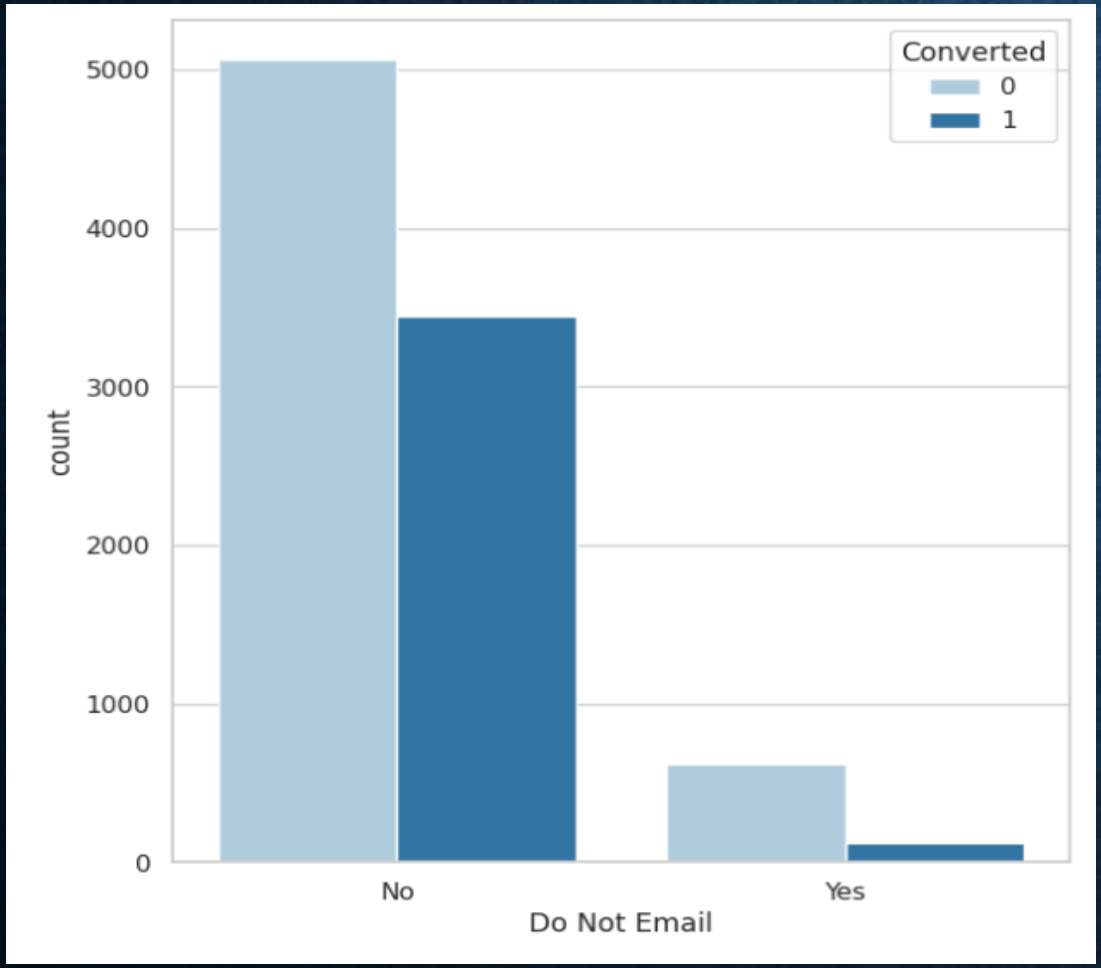
Landing page submissions has had high lead conversions.



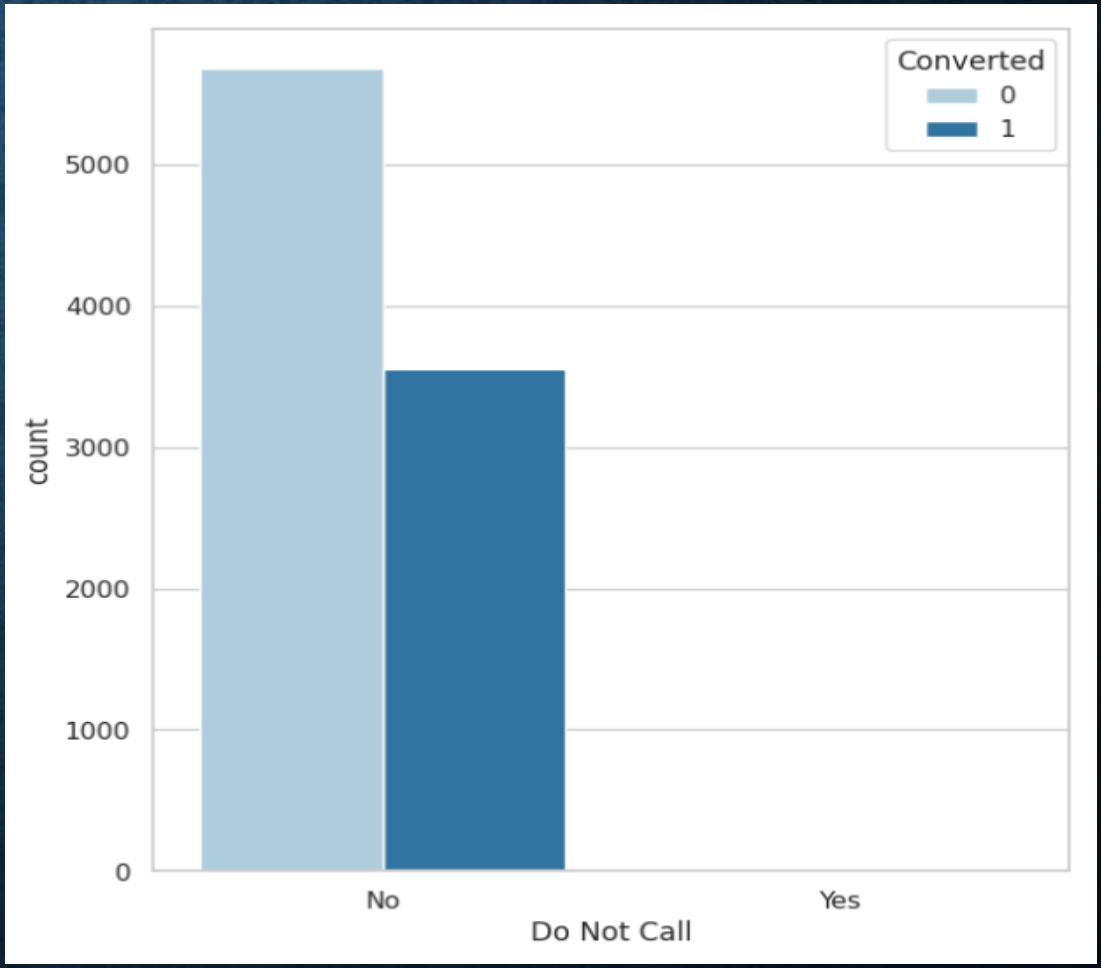
Google searches has had high conversions compared to other modes while references has had high conversion rate.



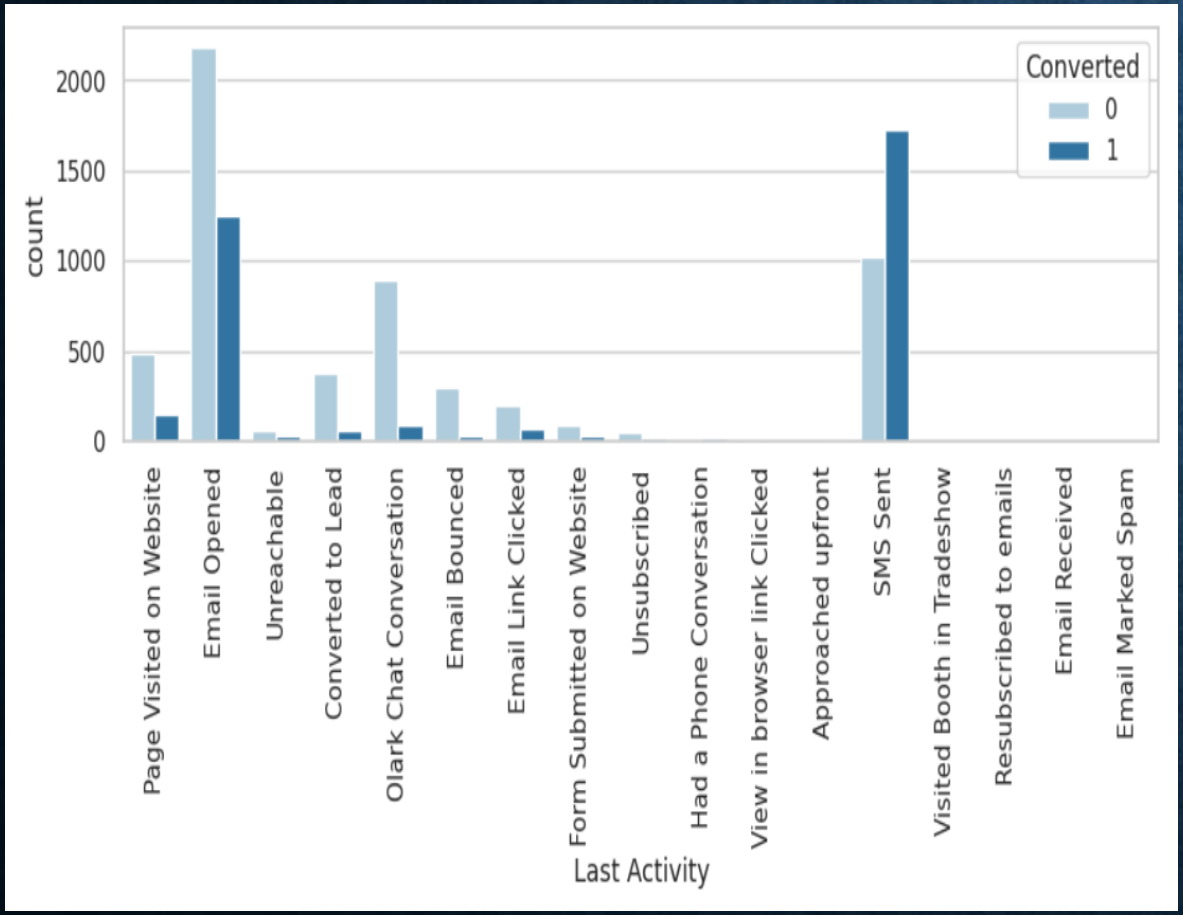
Most lead prefer not to informed through email.



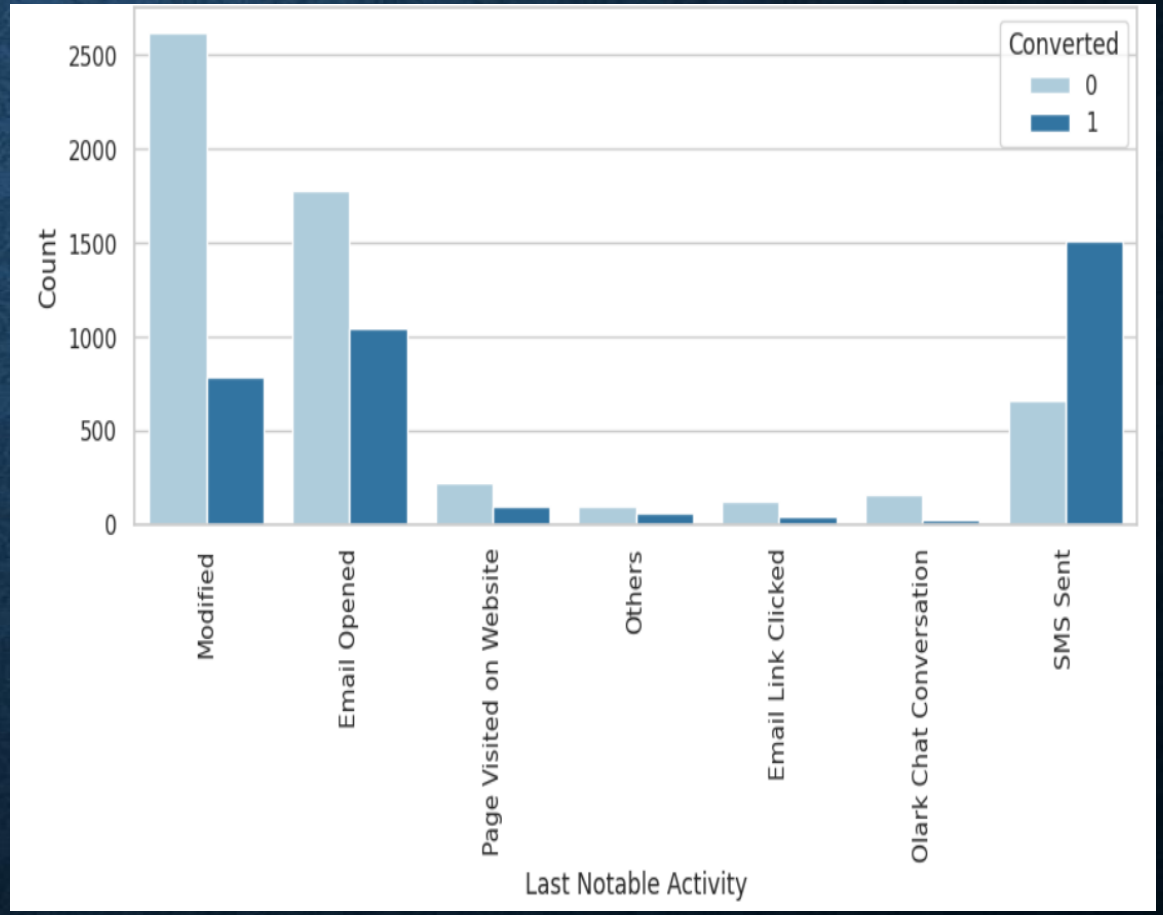
Most lead prefer not to informed through phone.



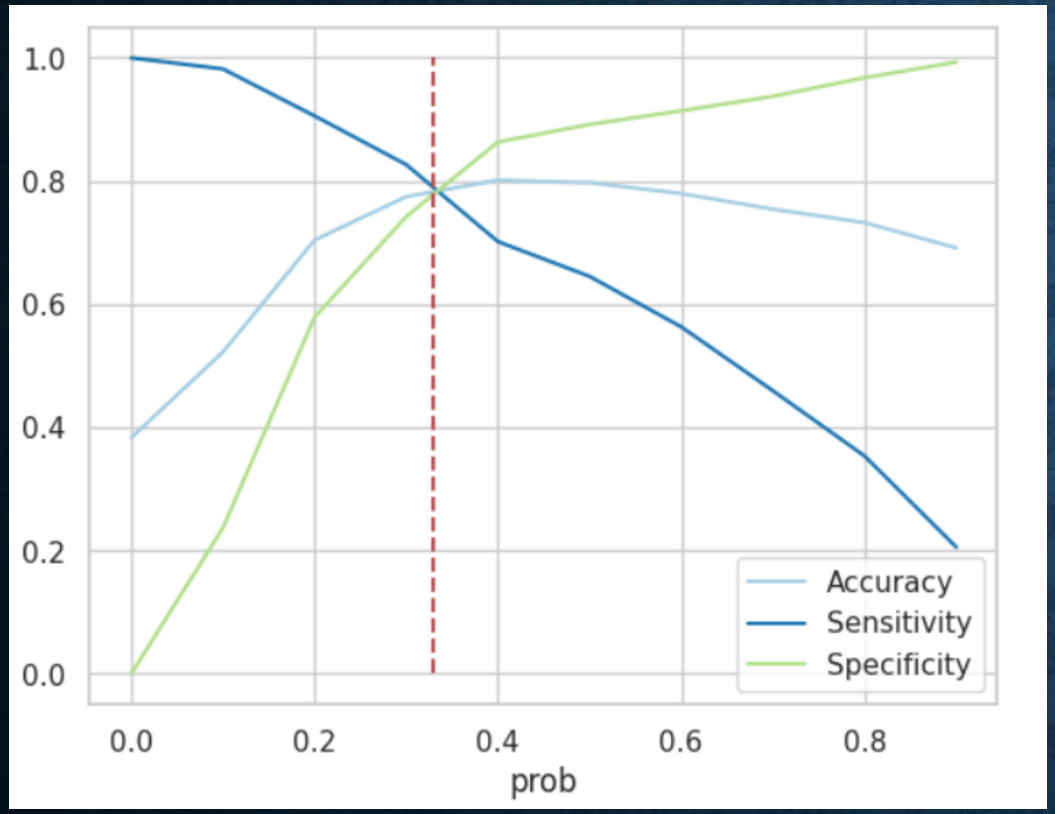
SMS has shown to be a promising method for getting higher confirmed leads, email also has high conversions.

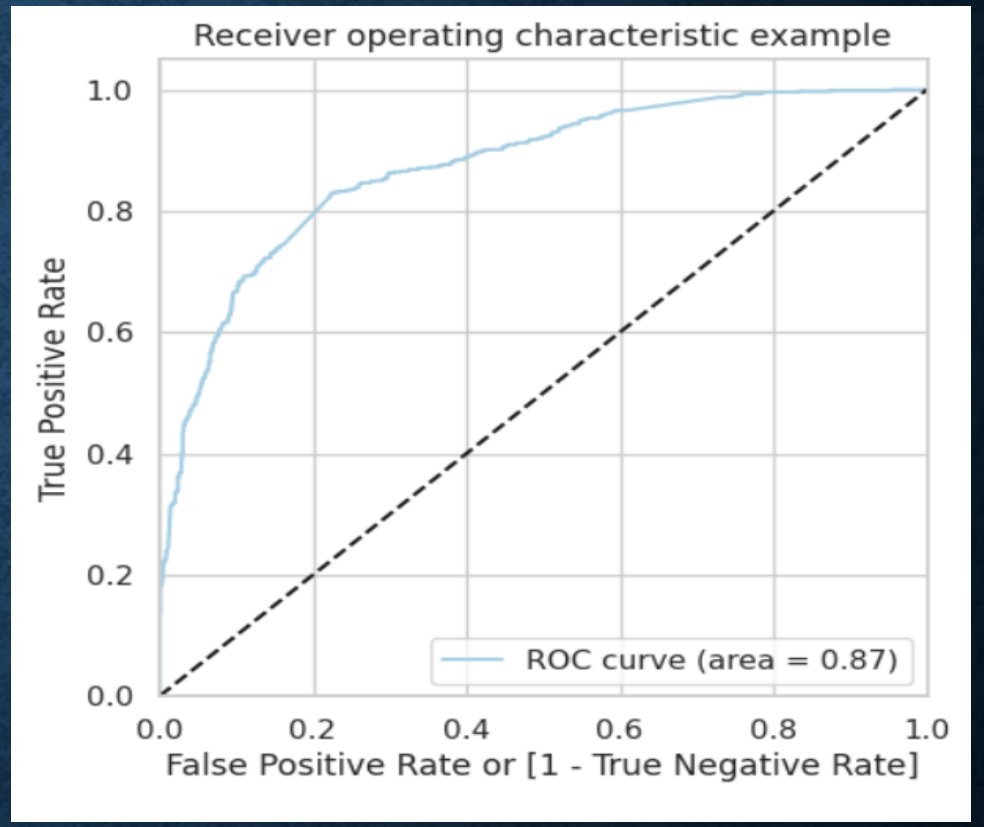


Most Lead are converted with messages, email also induce leads.



ROC CURVE





From the graph it is visible that the optimal cutoff is 0.33 where we have balanced sensitivity and specificity.

RESULTS FROM OUR ANALYSIS

Accuracy	= 0.8041
Sensitivity	= 0.738
Specificity	= 0.8472
False Positive Rate	= 0.1528
Precision	= 0.7588
Recall	= 0.738
Negative Predictive Value	= 0.8323

CONCLUSION

- People spending higher than average time are promising leads, so targeting them and approaching them can be helpful in conversion.
- Landing page submissions has had high lead conversions.
- SMS messages can have a high impact on lead conversion.
- References and offers for refering a lead can be good source for higher conversions.
- The model shows high close to 80% accuracy
- The threshould has been selected from Accuracy, Sensitivity, Specificity measures, precision and recall cuves.
- The model shows 73% sensitivity and 84% specificity.