Lead scoring case study

LOGISTIC REGRESSION

by Bhavani Ramachandran
Bharathi
Bishal Patel

An education company named X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google

x Education wants the most promising leads that can be converted to paying customers.

Although the company generates lots of leads only few are converted to paying customers.Wherein company wants a higher lead conversion.Leads comes through various enomorous mode like calls,mgs,emailgoogles search etc.

But the company has only 30% conversion rate through the whole process of turning leads into customers.The implementation process of lead generating attributes are not efficient in helping conversion.

Business Goal

The company needs a model to be built for selecting most promising leads

Lead score to be given to each leads.Higher the lead score are the Hot leads who are promising to be converted and lower the lead score are the cool lead.

Model to be built in conversion rate around 80%

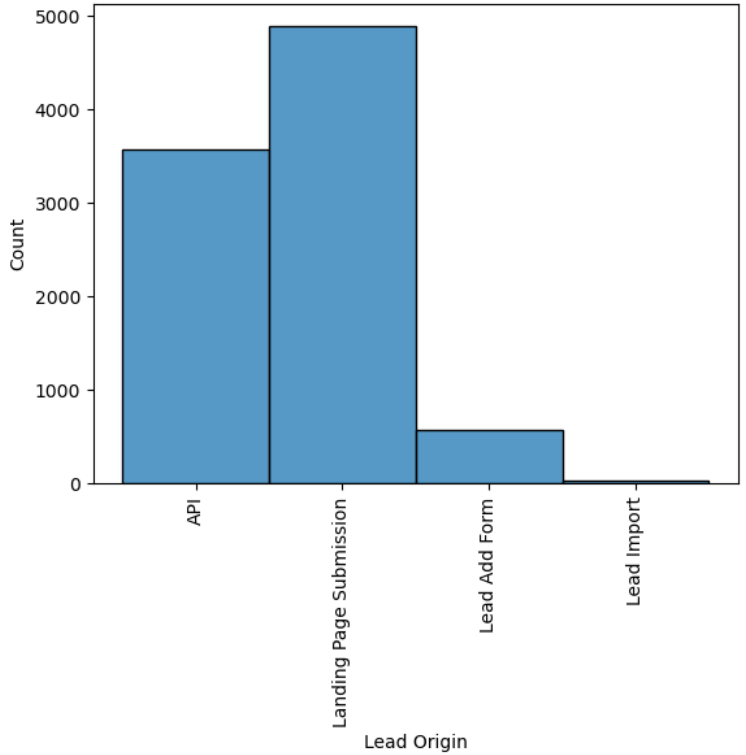
Strategy

Import the dataset
preprocessing for clean data
EDA for better analysis for conversion
scaling features
Prepare the data for Model building
Build a logistic regression model
Test the model on train data
Evaluate the model by different metrics
Test the model on test data
Measure the accuracy of the model and other evaluation metrics

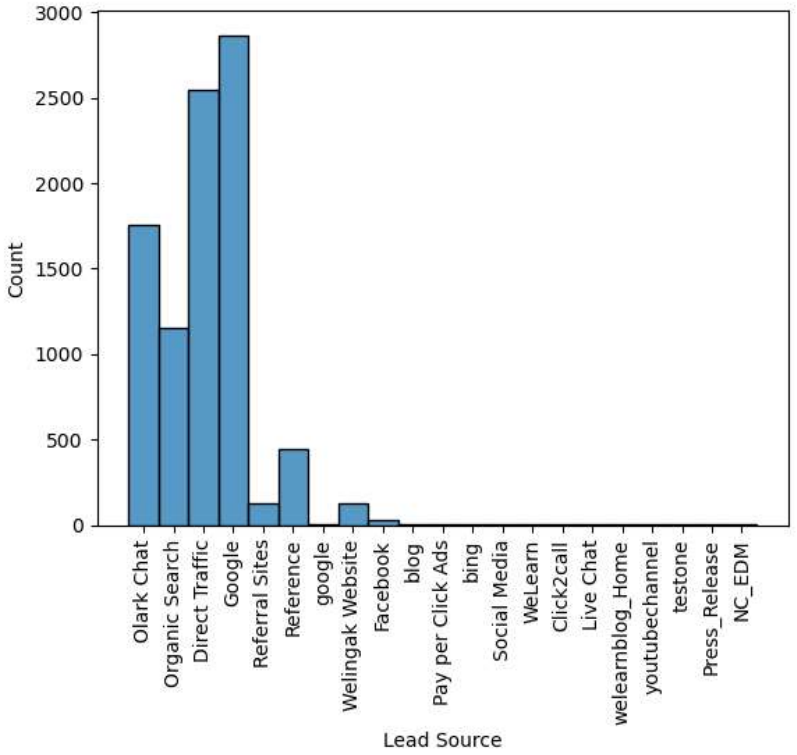
EDA Analysis

Categorical varaiable analysis

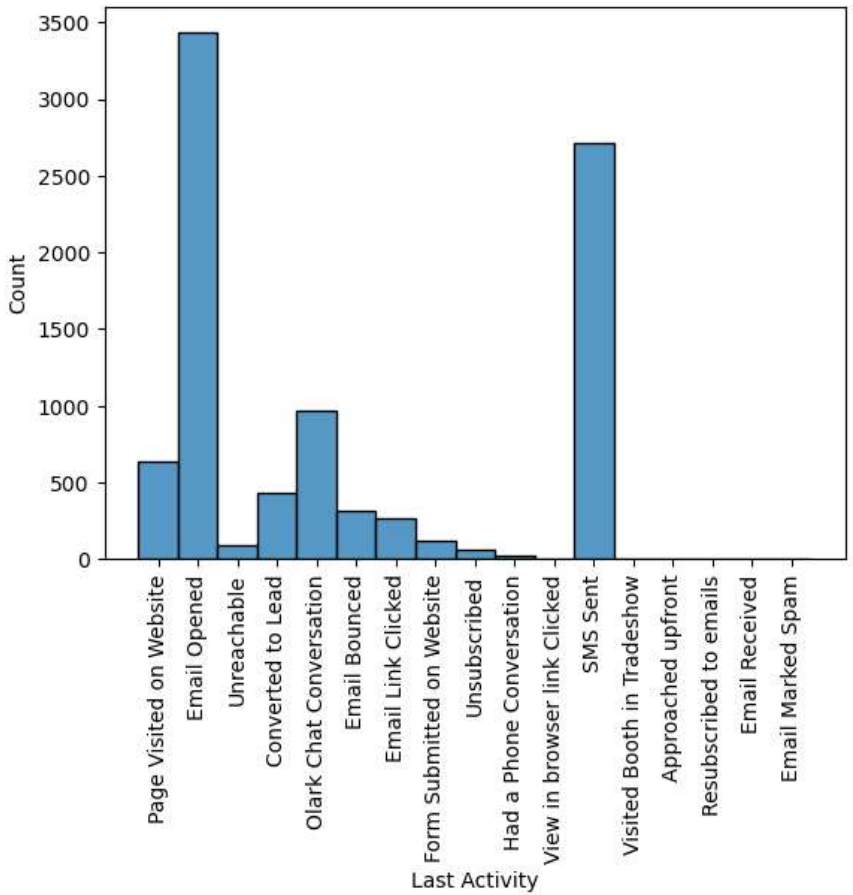
Here Landing page submission has high lead conversion

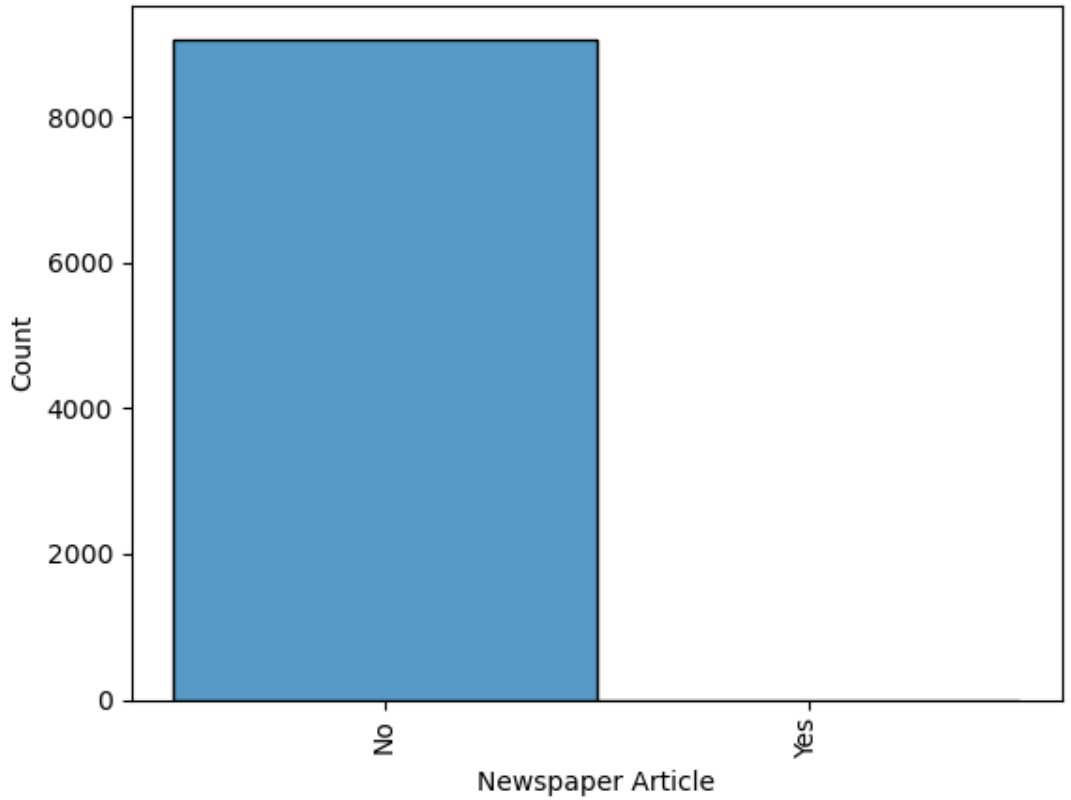


Olark chart,direct traffic and google has a high lead of conversion

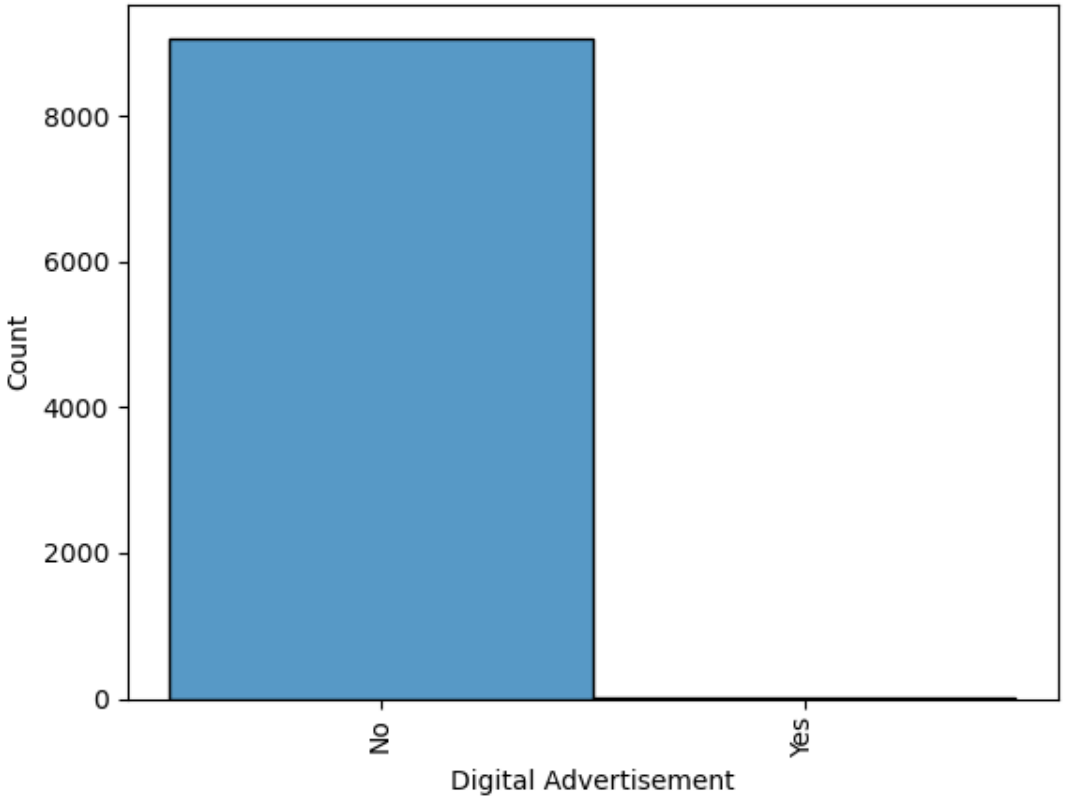


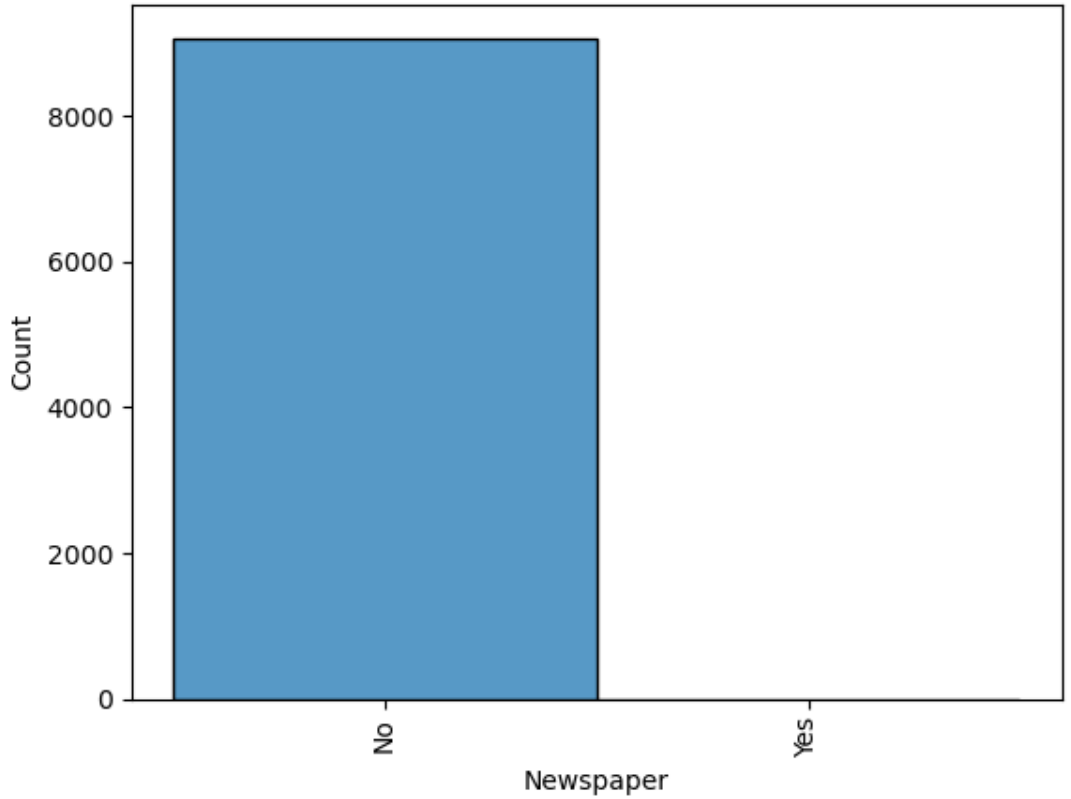
Here email opened,olark chat conversion and sms sent gives us more insights and this newspaper article does not provide a promising leaad



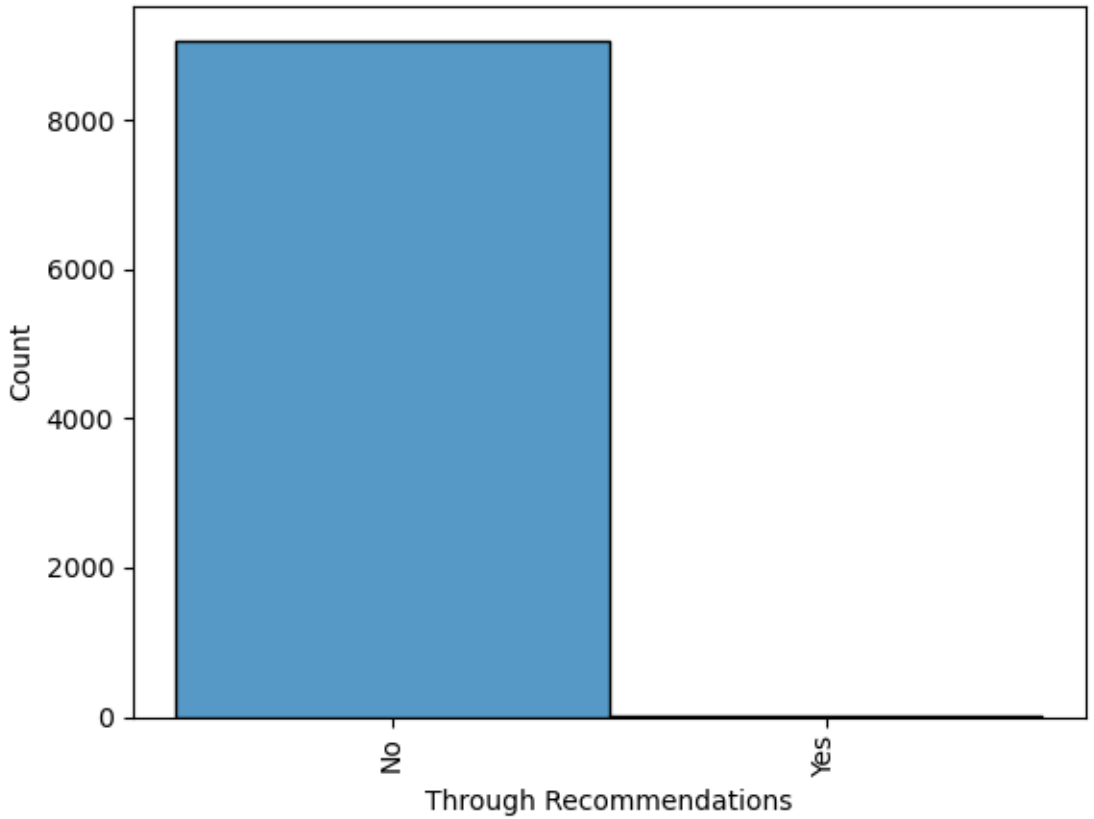


Digital advertisement and Newspaper are not the good source of lead

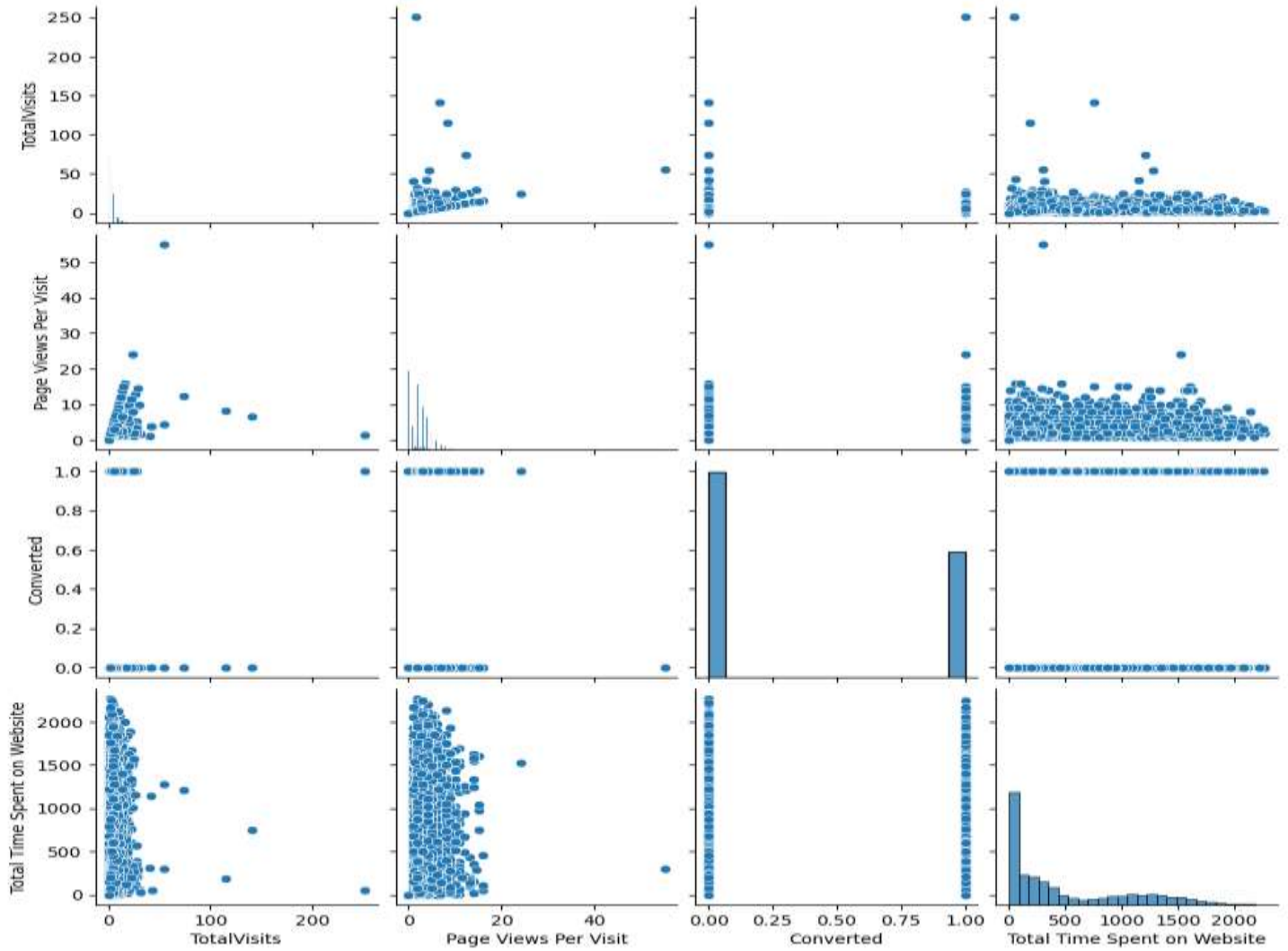




Through Recommendations also not good source of leads



People spending more time on website and page views per visit are higher chances of promising leads .

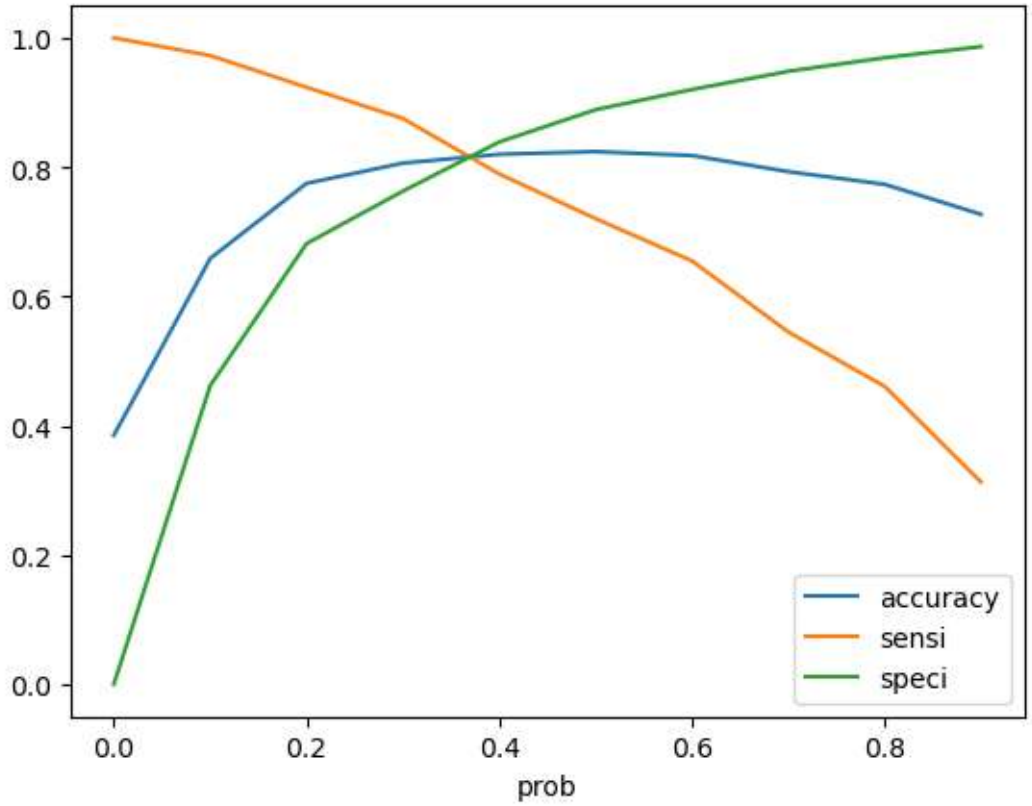


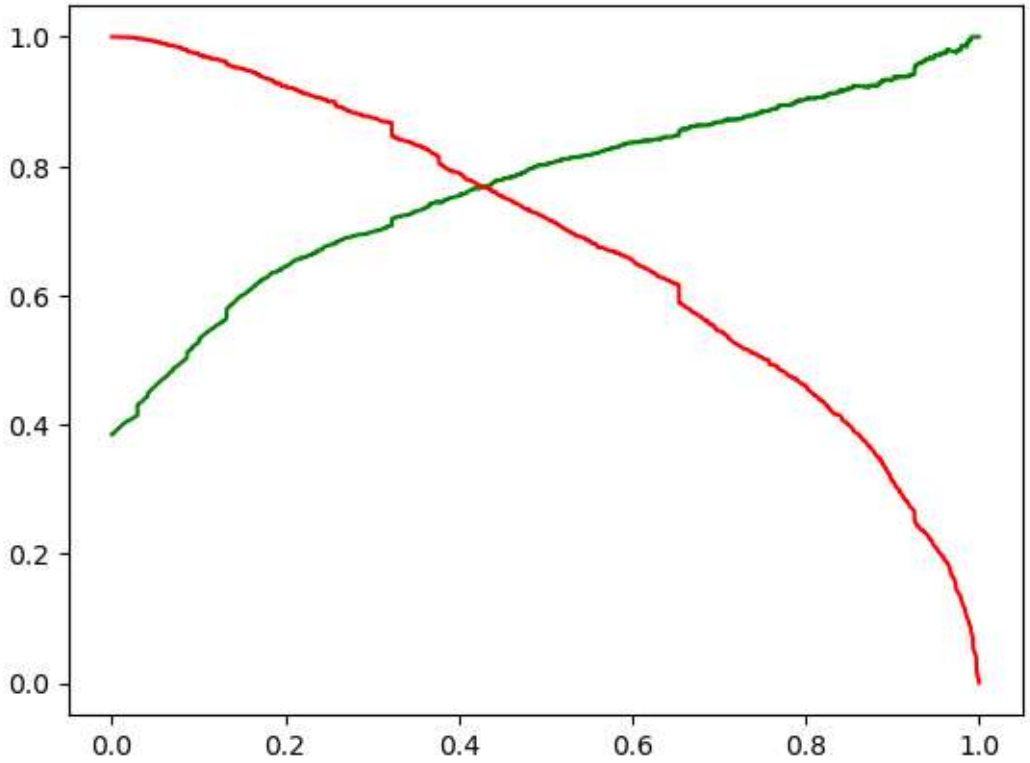
Model Building

- Splitting the train and test set
- Feature selection,here we have used standard scaler
- Build the first model
- Using the RFE and considering the p values eliminated the variables
- Now check for VIF values for all existing columns
- Predict using train set
- Evaluate the accuracy and other metrics
- Predict using test set
- Derived ROC curve
- Precision and Recall analysis on test predictions

Model Evaluation(Train set)

with accuracy of 80%
specificity 76%
sensitivity 87%
precision 80%
Recall 72%





Conclusion

People spending higher than average time on website are the having high chances of promising lead

SMS can have a high impact on lead conversion

Landing page submission can help out more leads

This model shows around 80% accuracy.

THANK YOU