LEAD SCORING CASE STUDY

BY
BALREDDYGUDA MOUNIKA
CH.MARUTHI SAI PHANI TEJA

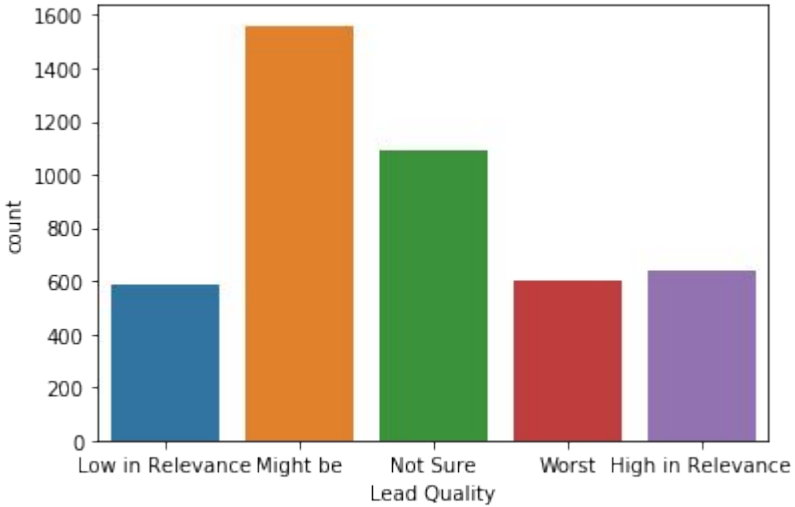
BRIEF OVERVIEW

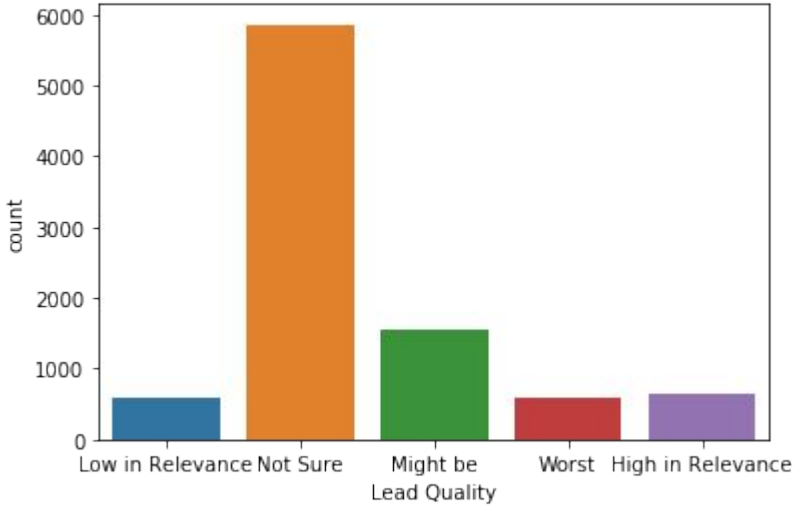
Building a model to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

STEPS TO BE PERFORMED

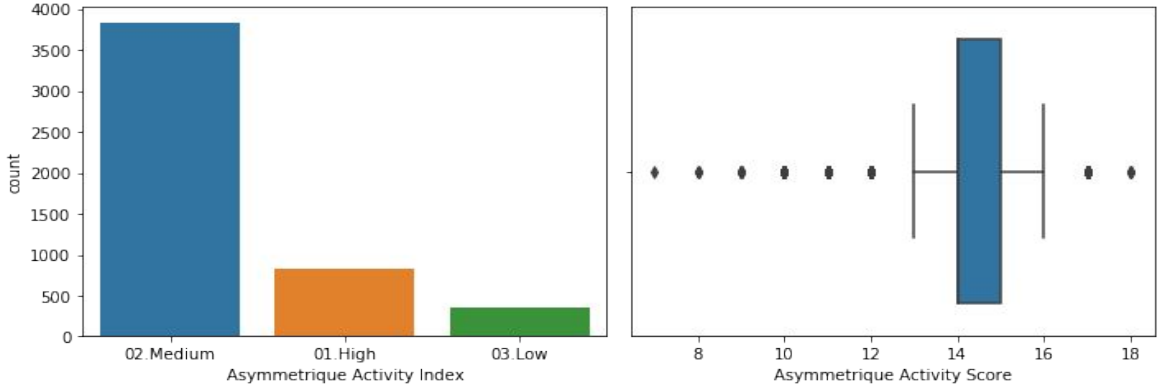
- Read and Visualise the data
- Data Cleaning
- Exploratory Data Analysis
- Data Preparation
- Test-Train Split
- Model building

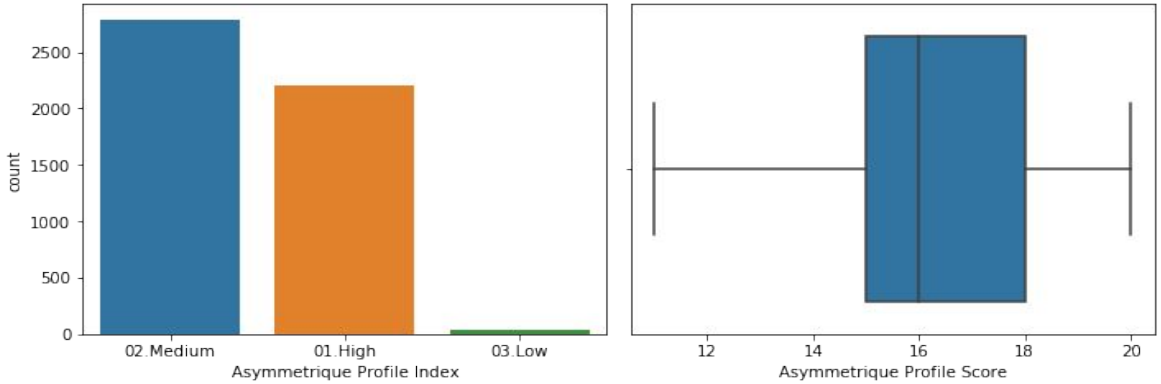
DATA CLEANING





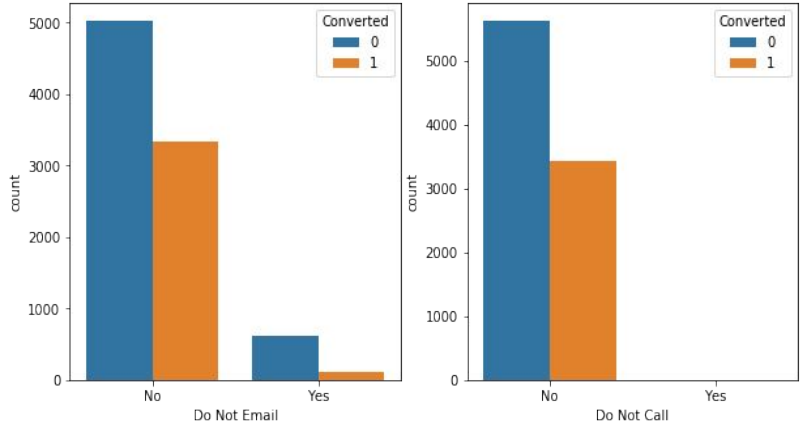
As Lead quality is based on the intuition of employee,analyze the column and fill the null_values through the percentage of the sub-categories in the Lead_quality

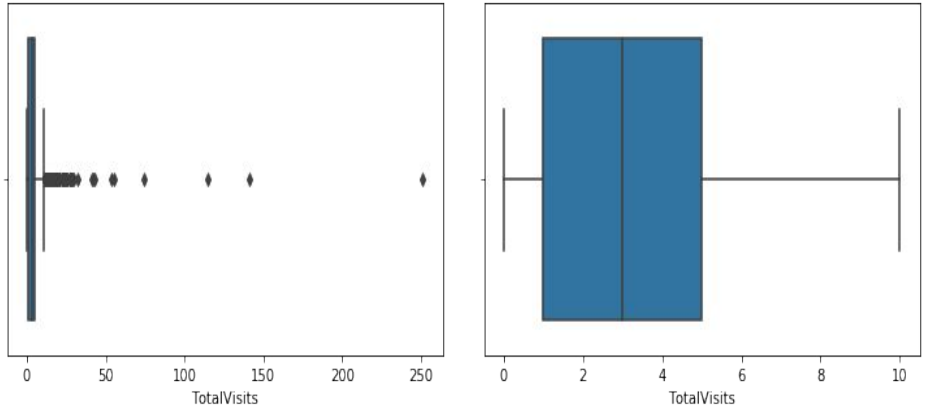




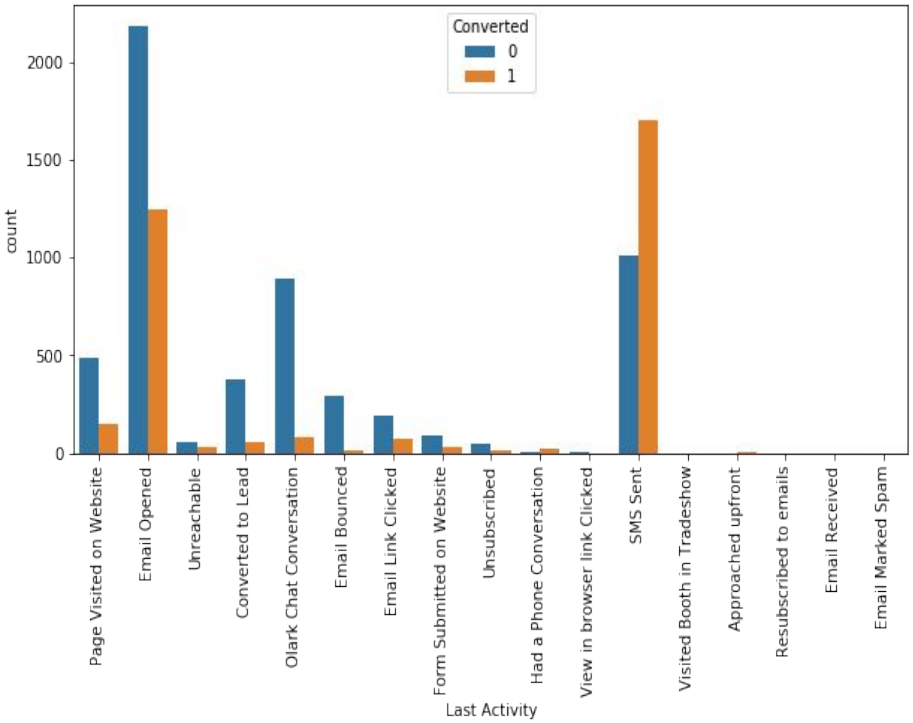
There is too much variation in these parameters so it's not reliable to impute any value in it so we need to drop these columns

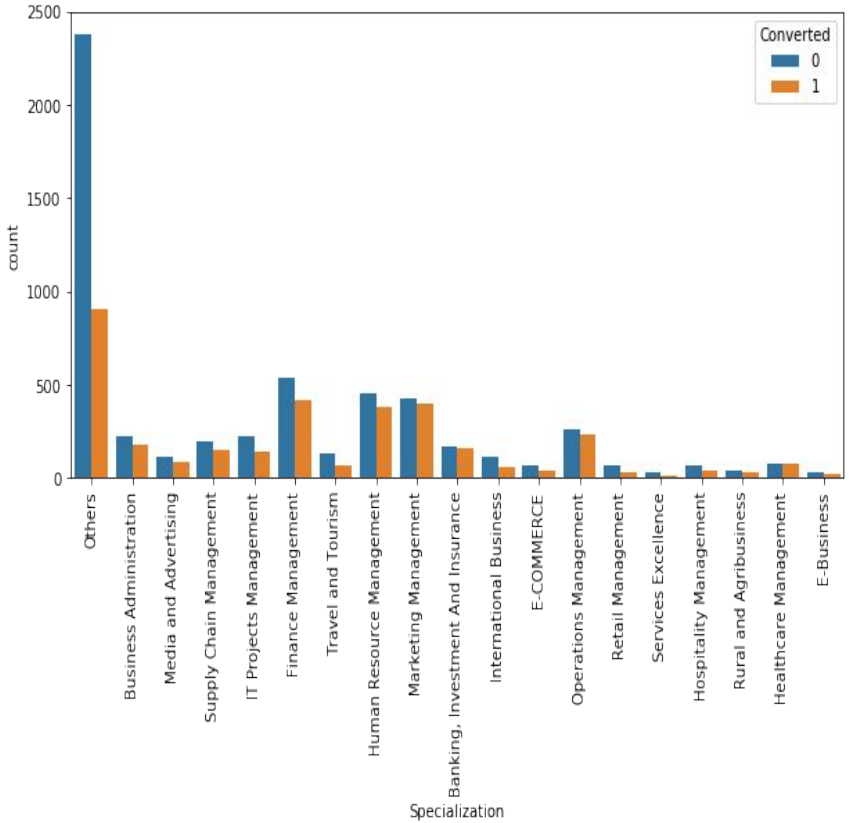
EXPLORATORY DATA ANALYSIS



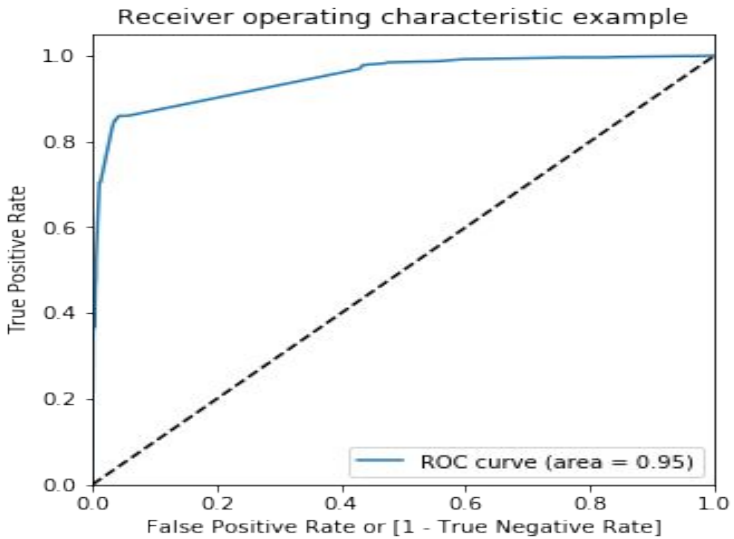


As we can see there are a number of outliers in the data we will cap the outliers to 95% value for analysis.

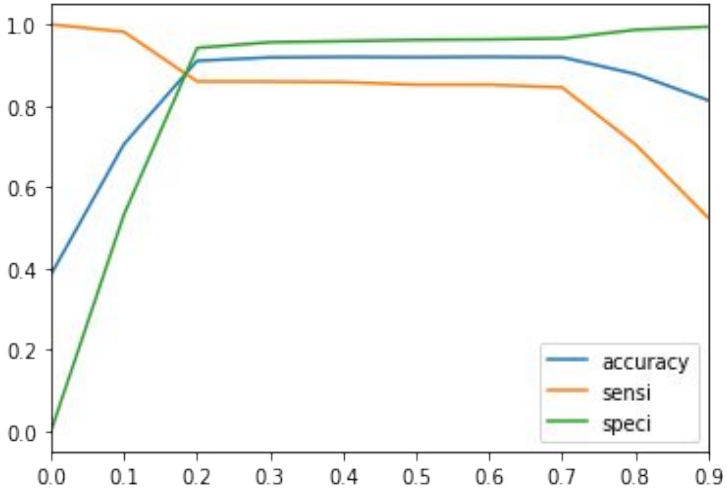

Converted
0
1
count
2000
1500
1000
500
0
Page Visited on Website
Email Opened
Unreachable
Converted to Lead
Olark Chat Conversation
Email Bounced
Email Link Clicked
Form Submitted on Website
Unsubscribed
Had a Phone Conversation
View in browser link Clicked
SMS Sent
Visited Booth in Tradeshow
Approached upfront
Resubscribed to emails
Email Received
Email Marked Spam
Last Activity


Converted
0
1
count
2500
2000
1500
1000
500
0
Others
Business Administration
Media and Advertising
Supply Chain Management
IT Projects Management
Finance Management
Travel and Tourism
Human Resource Management
Marketing Management
Banking, Investment And Insurance
International Business
E-COMMERCE
Operations Management
Retail Management
Services Excellence
Hospitality Management
Rural and Agribusiness
Healthcare Management
E-Business
Specialization

PLOTTING ROC CURVE

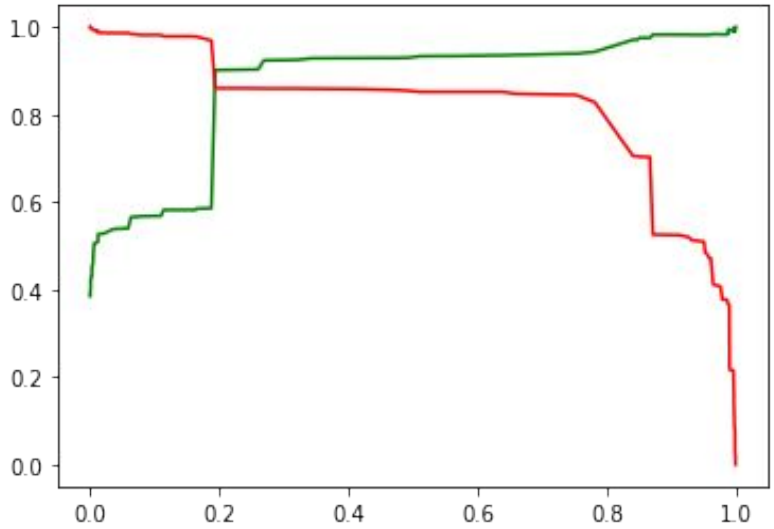


FINDING OPTIMAL CUTOFF



From the curve above, 0.2 is the optimum point to take it as a cutoff probability.

PRECISION AND RECALL TRADEOFF



OVERALL METRICS FOR LOGISTIC REGRESSION MODEL

Accuracy = 0.90

Sensitivity = 0.84

Specificity = 0.94