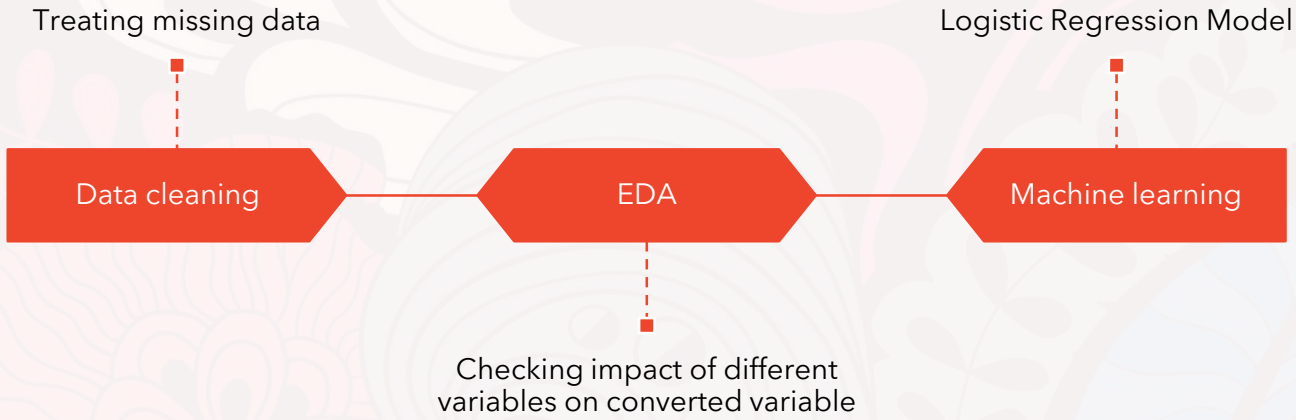
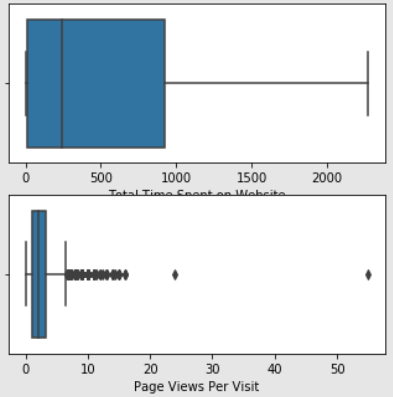
MOST PROMISING LEADS

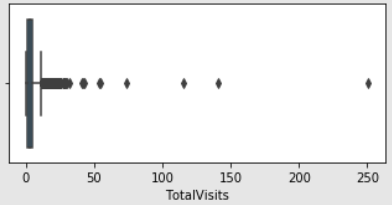
Techniques used

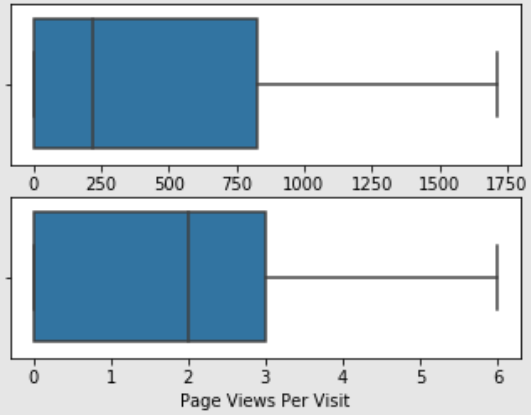


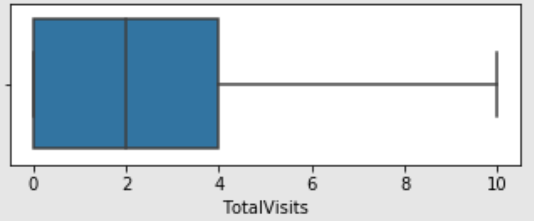
Data Cleaning

- Firstly we have removed the columns which had more than 40% missing values
- Secondly we have Treated the missing values by using the methods like imputation on columns with missing values less than 40%
- We have also did the outlier treatment on continuous columns.(below images shows the same)



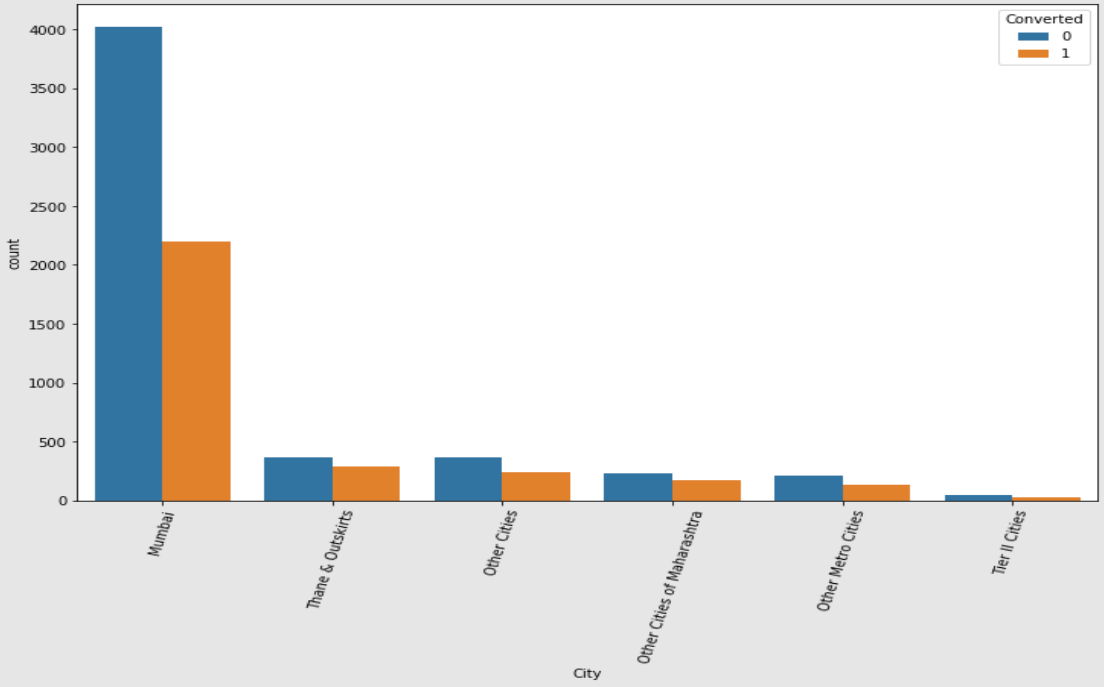


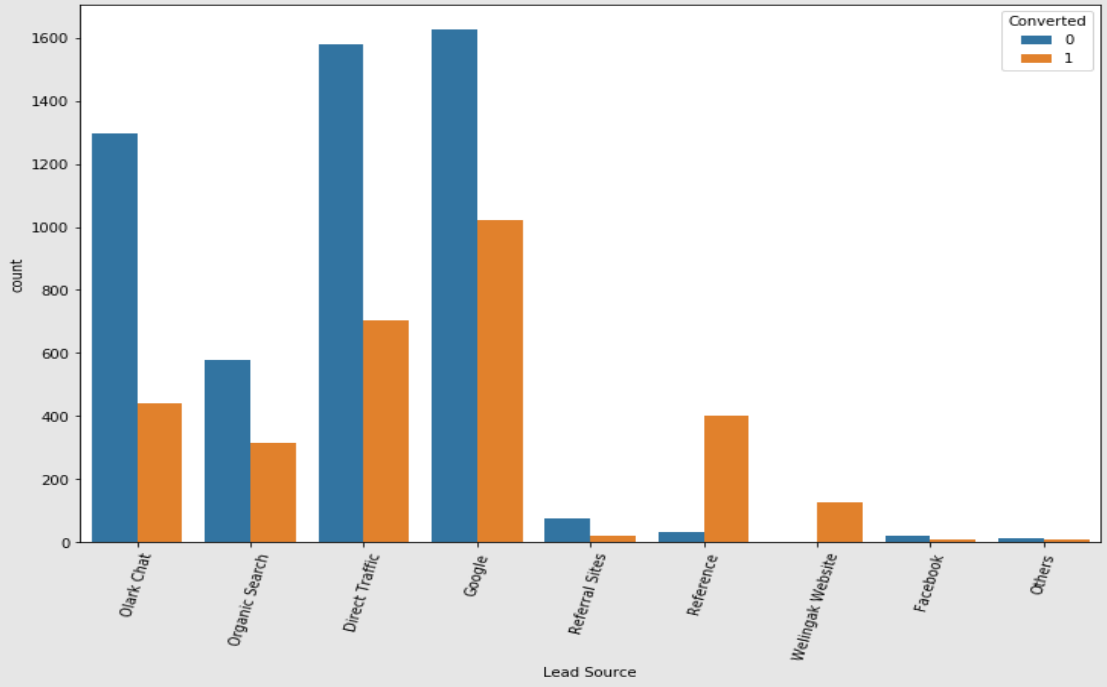




EDA done on variables wrt Coverted column

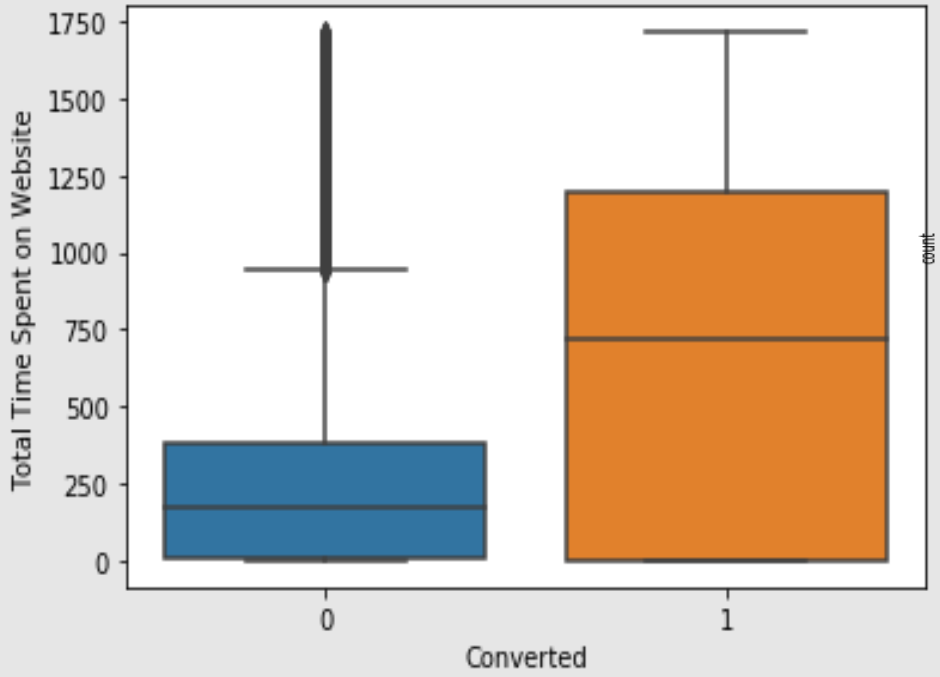
- Approx 30% leads are converted in Mumbai.
- Lead Source conversion rate : Olark chart- 25% ,Direct traffic :30%, Google : 40% also we can see the sourceswith less leads like Reference and Welingak have a great conversion rate of more than approx 70%.

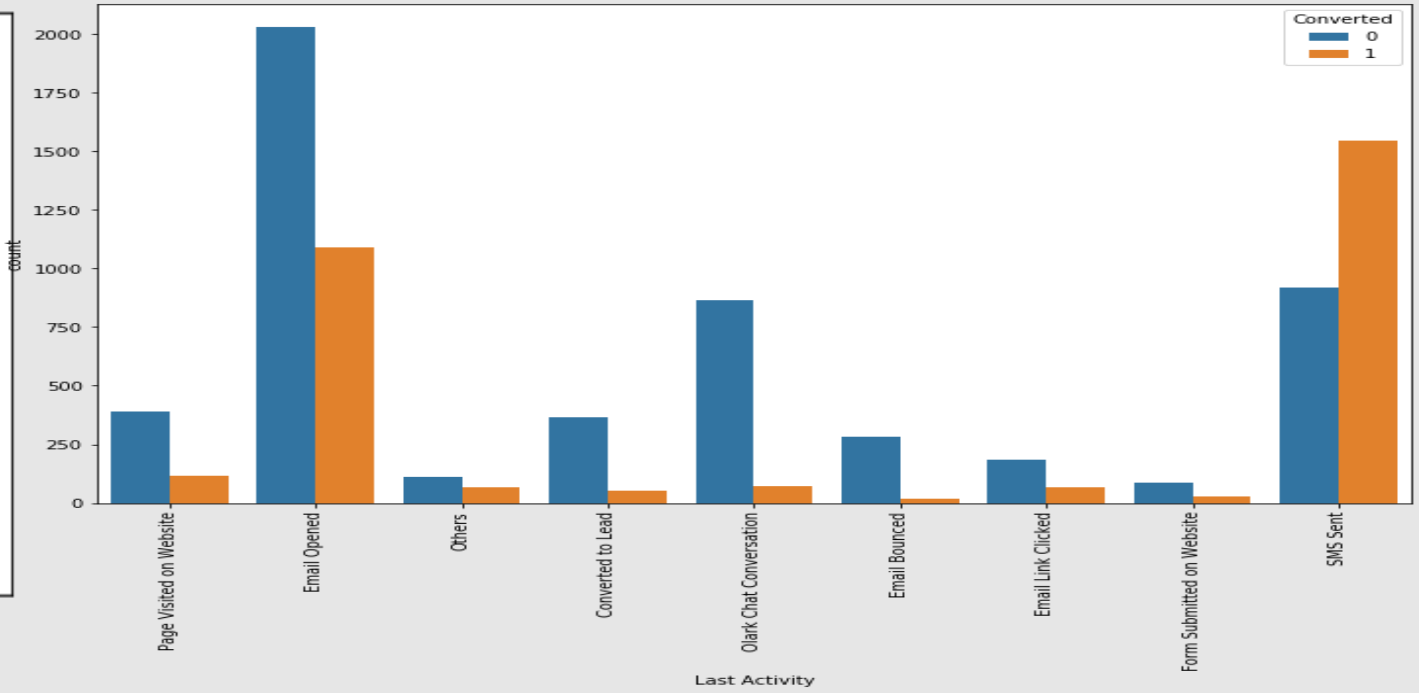




EDA done on variables wrt Coverted column(cont)

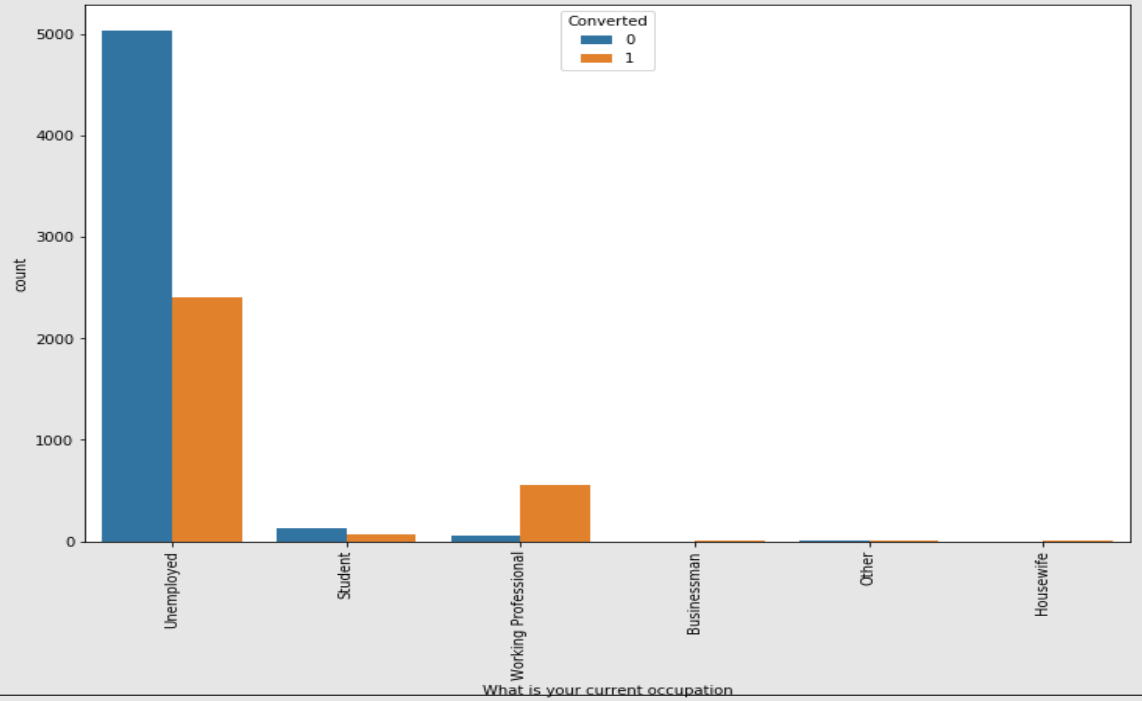
- Leads who converted spent 3x time more on website compared to not converted once..
- Last activity of leads with conversion rate : opened the email 33% and sent sms with approx. 60%



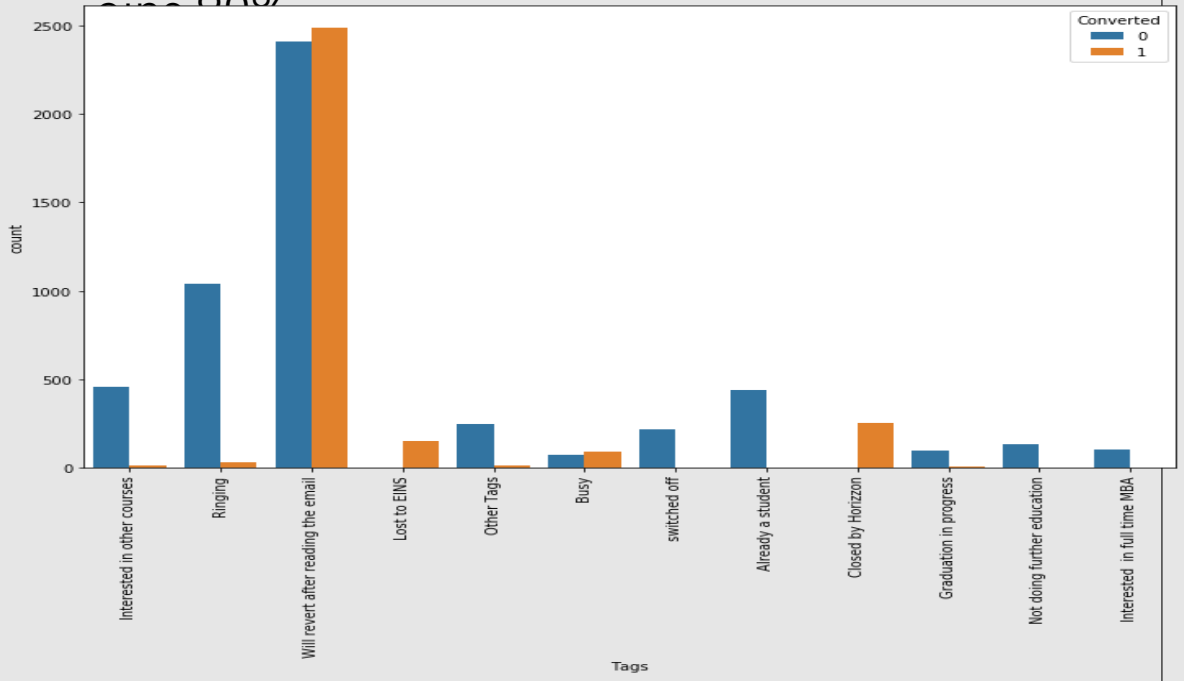


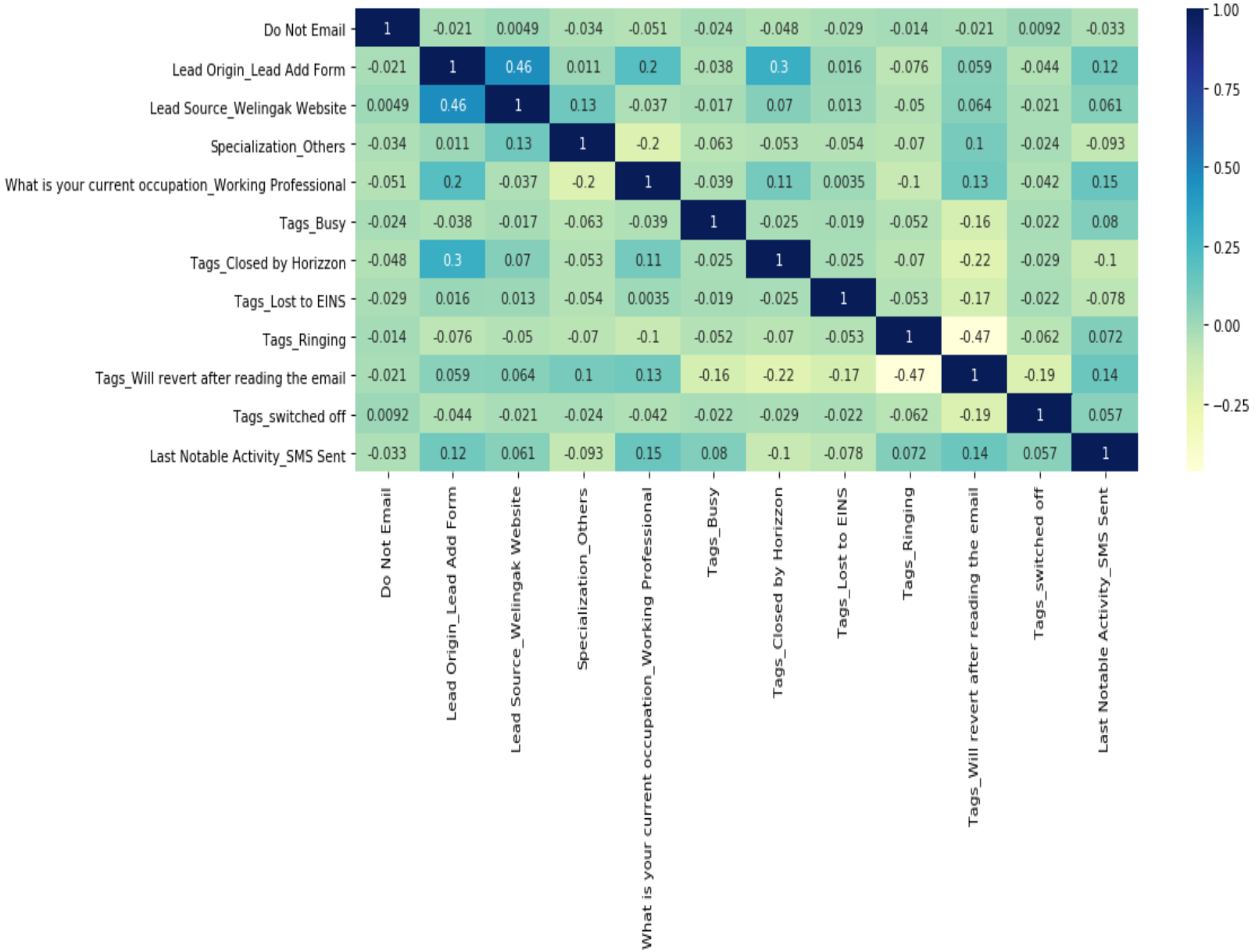
EDA done on variables wrt Coverted column(cont)

- Current occupation conversion rate : unemployed : 33% and working professional : 70%



- Tags of the leads with conversion rate : revert after reading email :more than 50% , closed by horizon 90% and lost to





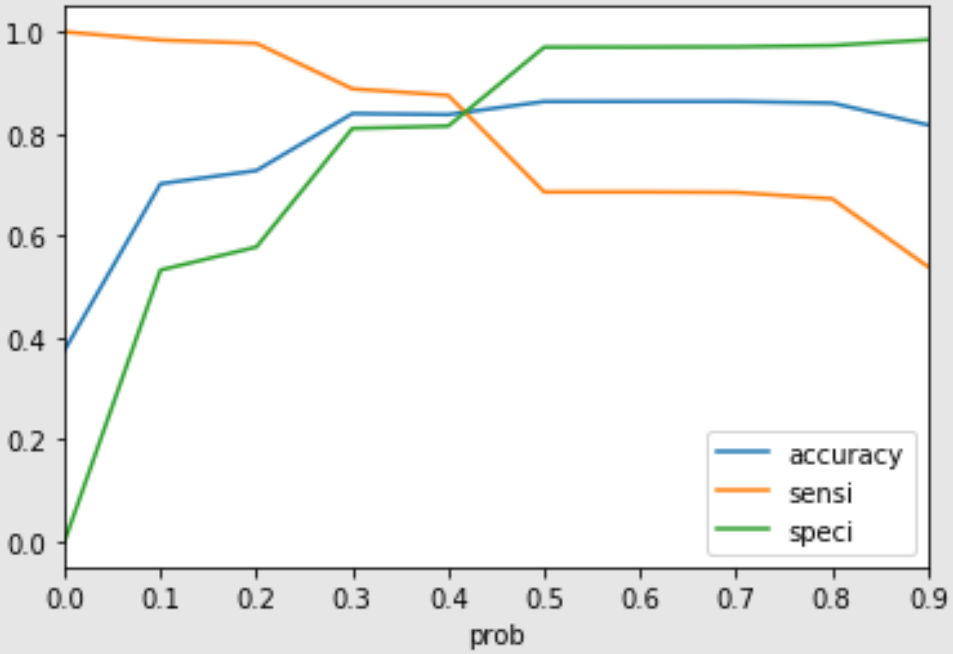
Correlation of top significant variables choosed by RFE technique

Optimal Cut off : 0.4

Results :

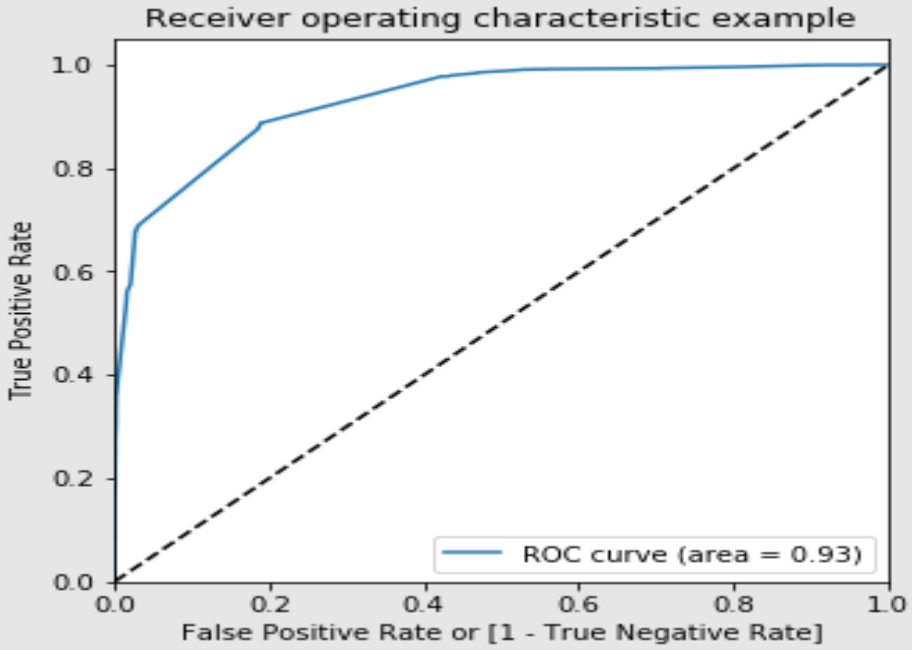
Data	Train	Test
Accuracy –	83.8	83.1
Sensitivity-	87.5	86.9
Specificity-	81	81
Precision –	74	71.4
Recall -	87.5	86.9
F1 -	80.1	78.1

Optimal cut off graph :

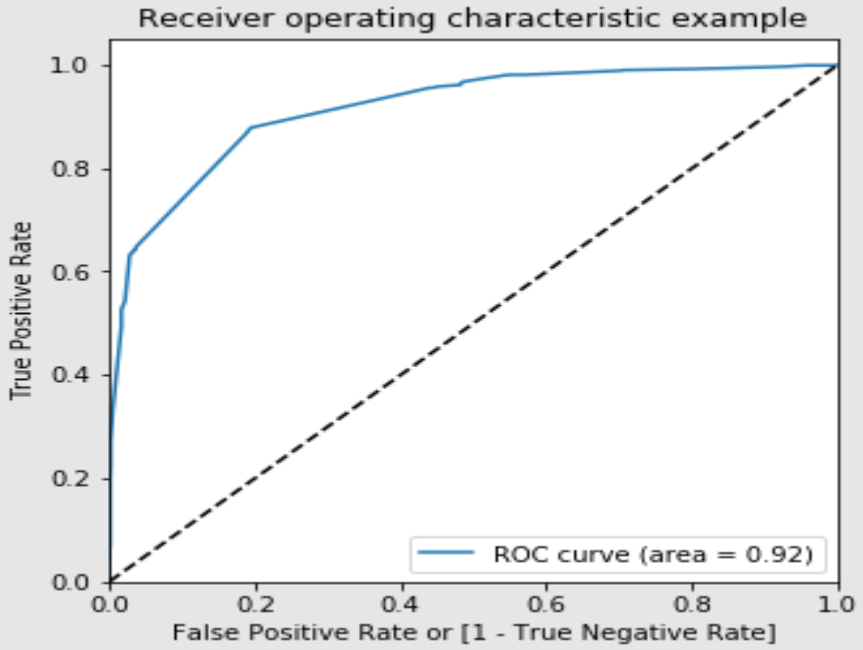


ROC Curve

Train Data



Test Data



According to the model lead score following are the Hot leads who will convert in order

THANK YOU