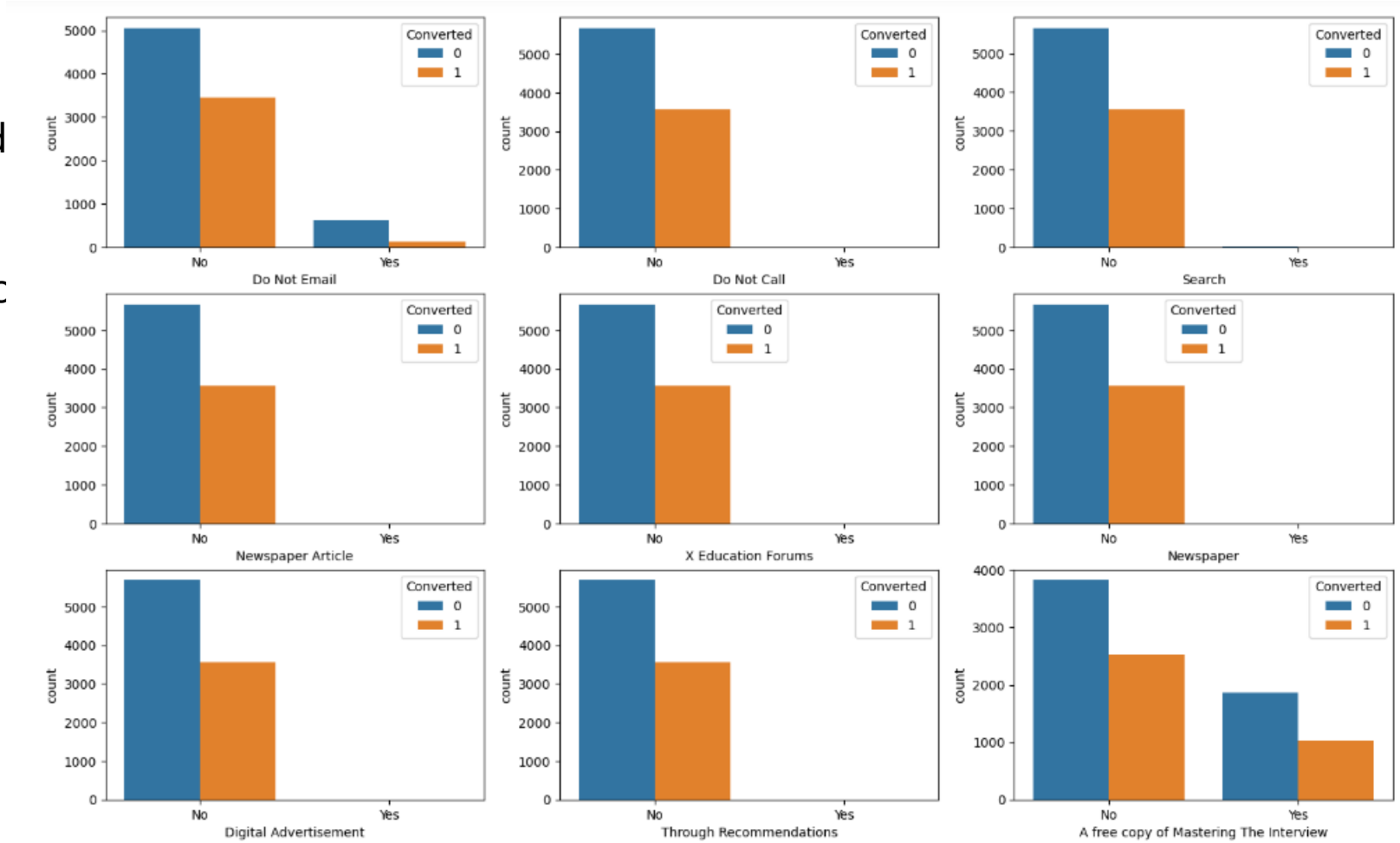
Lead Scoring Case Study

Sonia Lulla
Harshal Ajmera
Vidisha Bhosle

EDA & Outliers Treatment: Bivariate Analysis

1

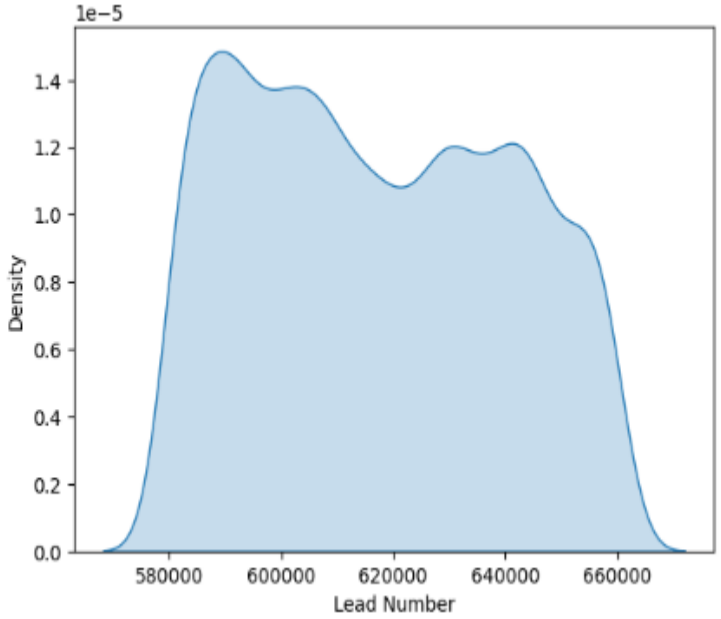
- Conversions is higher for those who inquired through digital marketing
- Conversion is higher fc those who asked Not Email, Not to call

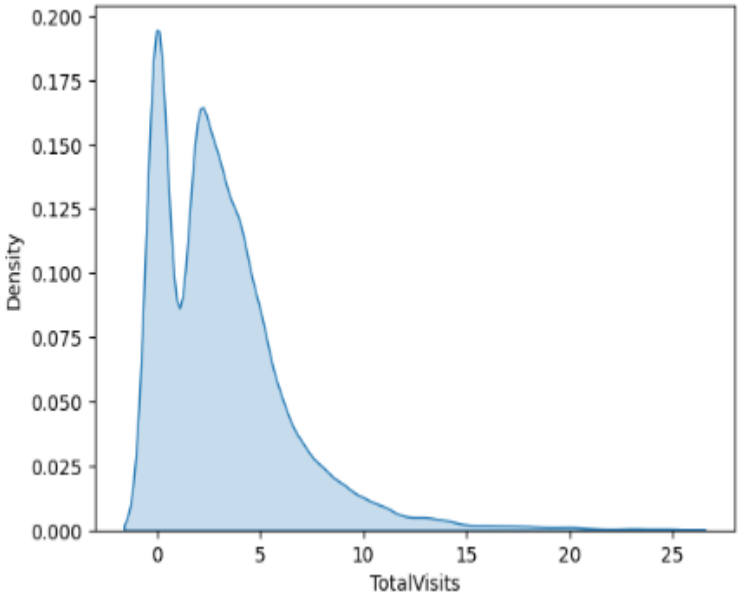


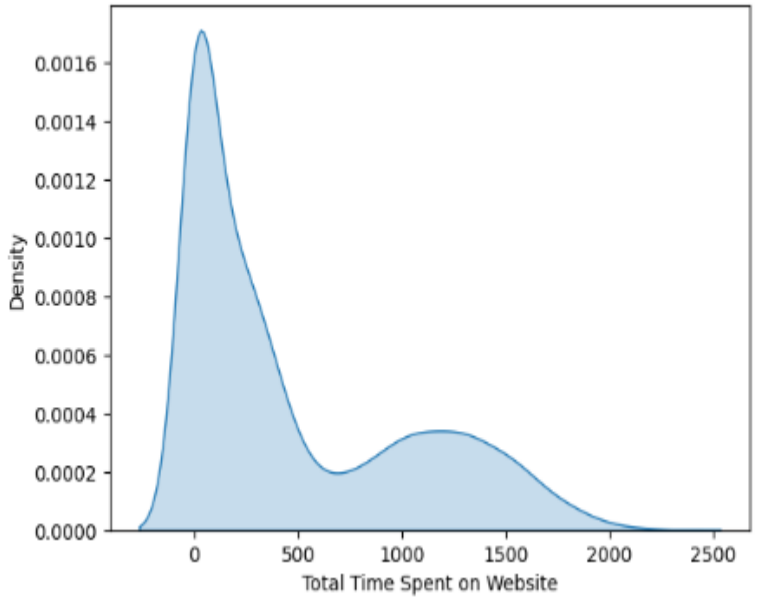
Observation of Bivariate Analysis on Numerical Variables:

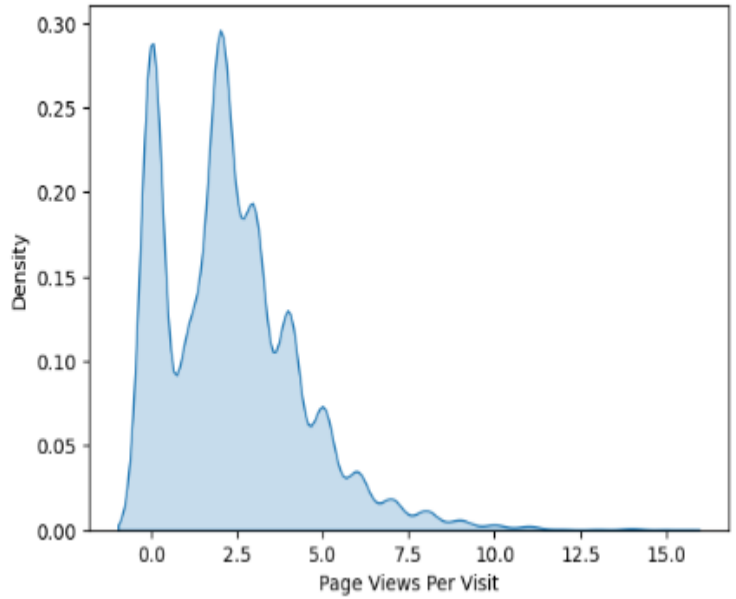
Conversion is higher with below pointers:

1. The number of visits to website is more
2. The total Time spent on website is more
3. The number of pages visited is more

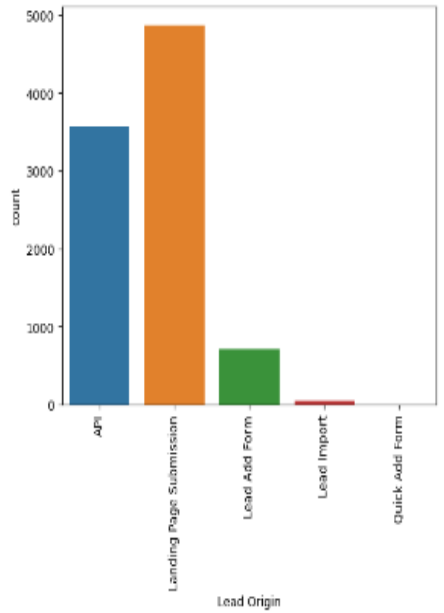


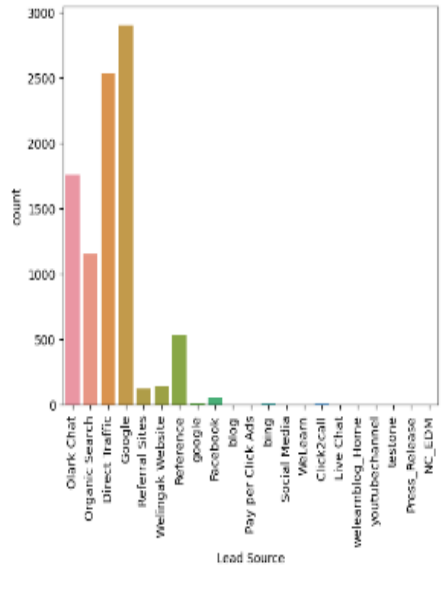


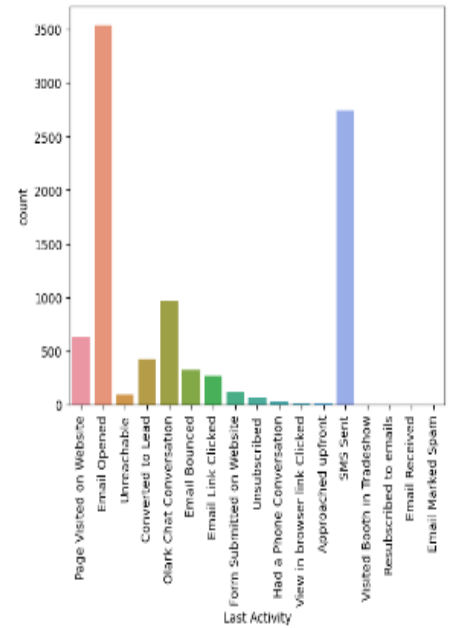


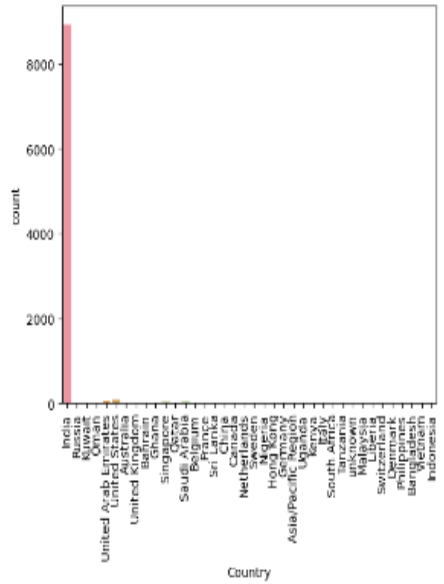


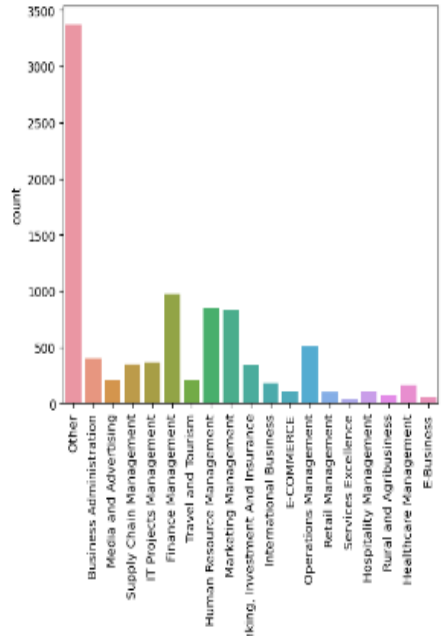
- Most of the leads are coming from landing page submission
- Source of most of the leads is Google, Direct Traffic
- Most of the visitors are unemployed
- Most of them are looking at course due to better career perspectives
- Highest conversion is from India

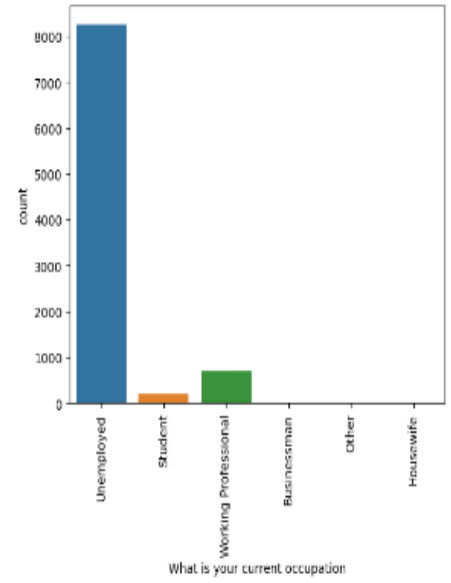


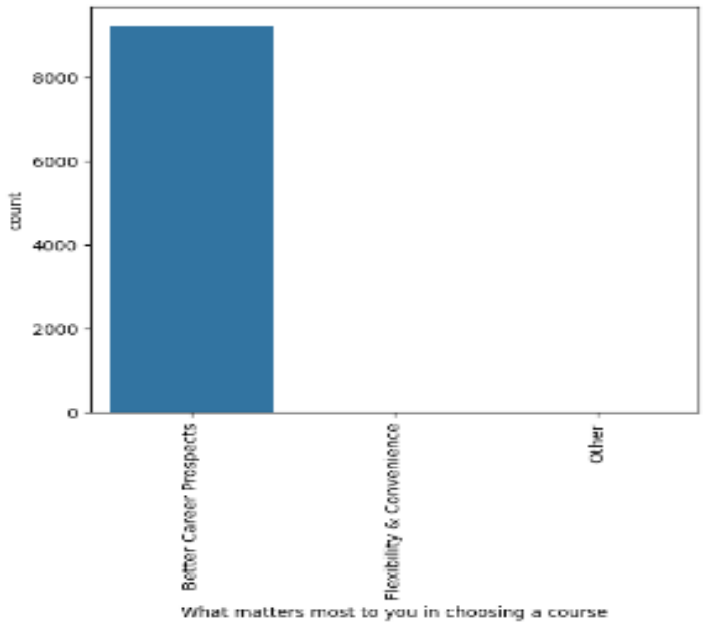


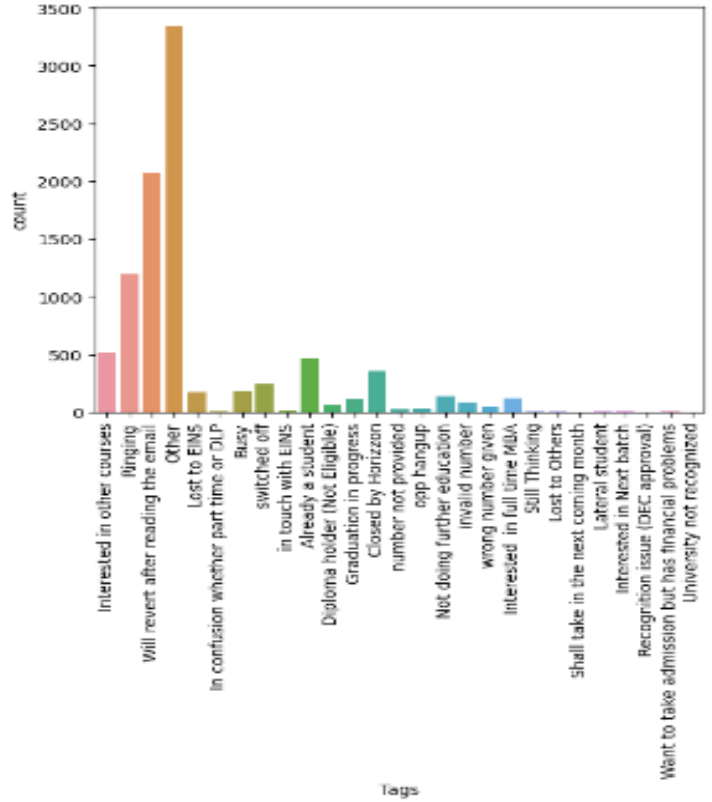


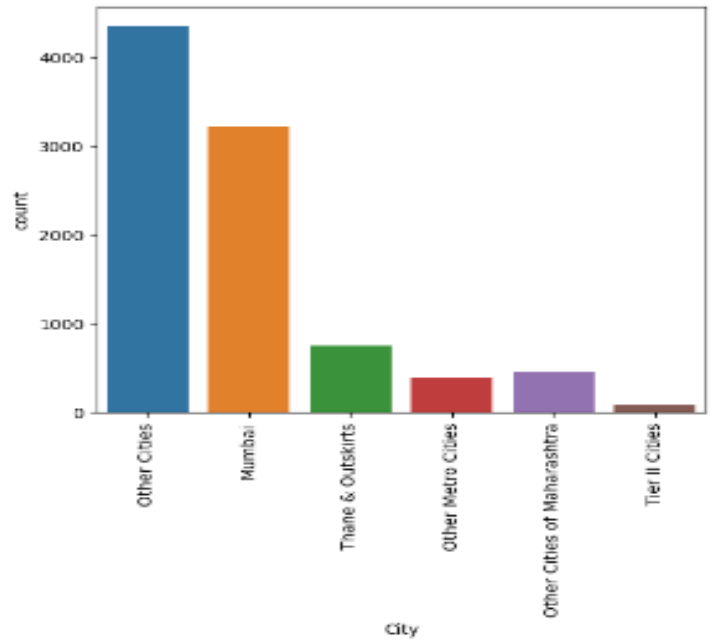


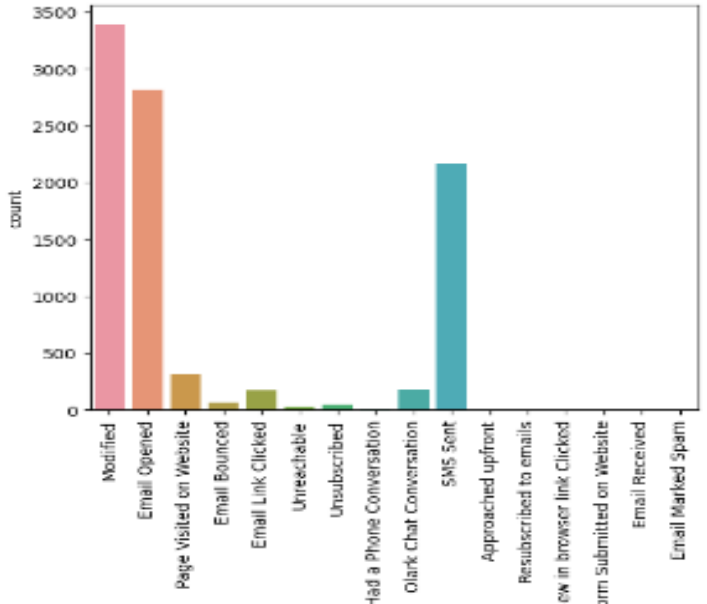


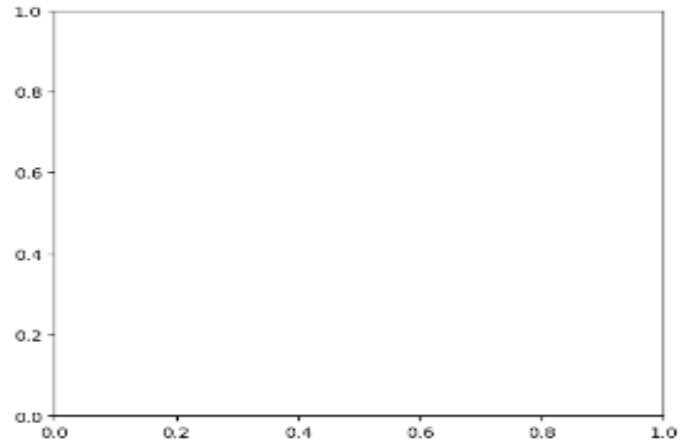


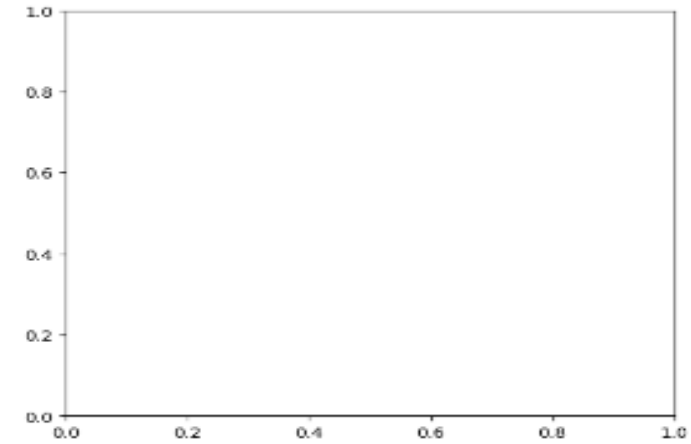

count
8000
6000
4000
2000
0
Better Career Prospects
Flexibility & Convenience
Other
What matters most to you in choosing a course


count
3500
3000
2500
2000
1500
1000
500
0
Interested in other courses
Ringing
Will revert after reading the email
Other
Lost to EINS
In confusion whether part time or DLP
Busy
switched off
in touch with EINS
Already a student
Diploma holder (Not Eligible)
Graduation in progress
Closed by Horizzon
number not provided
opp hangup
Not doing further education
invalid number
wrong number given
Interested in full time MBA
Still Thinking
Lost to Others
shall take in the next coming month
Lateral student
Interested in Next batch
Recognition issue (DEC approval)
Went to take admission but has financial problems
University not recognized
Tags


count
4000
3000
2000
1000
0
Other Cities
Mumbai
Thane & Outskirts
Other Metro Cities
Other Cities of Maharashtra
Tier II Cities
City


count
3500
3000
2500
2000
1500
1000
500
0
Modified
Email Opened
Page Visited on Website
Email Bounced
Email Link Clicked
Unreachable
Unsubscribed
Had a Phone Conversation
Olark Chat Conversation
SMS Sent
Approached upfront
Resubscribed to emails
Email Received
Email Marked Spam

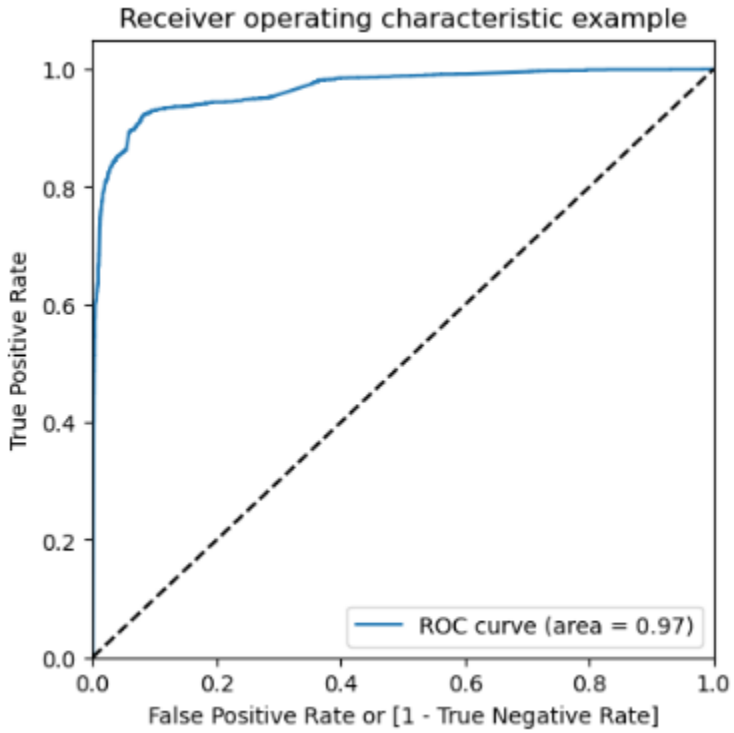

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

CONCLUSION FROM EDA:

1. Total time spent on website
2. Lead Source
3. Last Notable Activity
4. TAGS

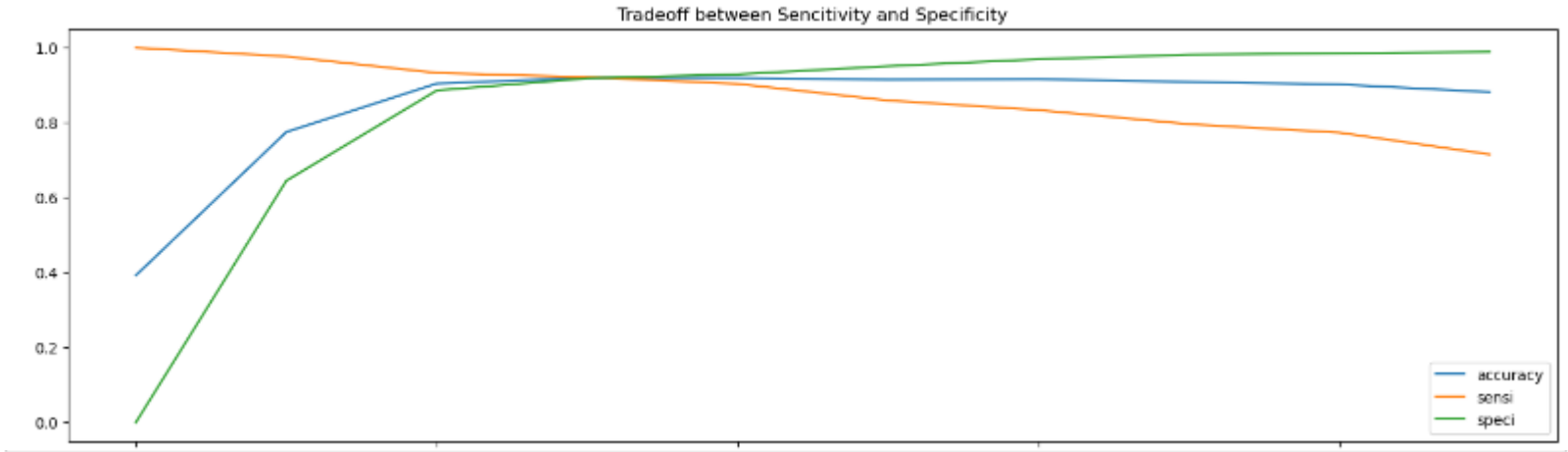
Logistic Regression Model

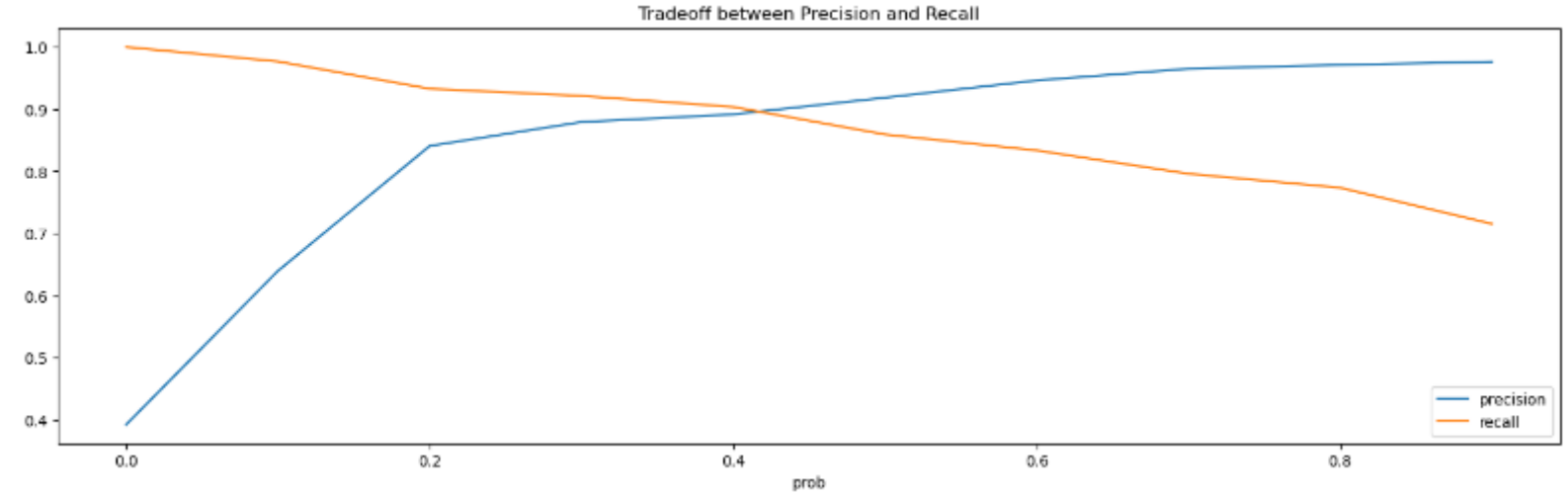


- **AREA UNDER CURVE IS 0.97**
- That means the model is highly capable of distinguishing between classes.

Determining Sensitivity Of the Model:

<Figure size 700x500 with 0 Axes>





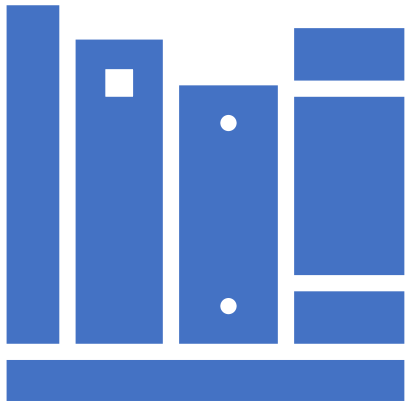
By observing the tradeoff between sensitivity and specificity we can conclude that the cut off an be 27 %

Model Summary:

- **most important features are**

1. Tags_Closed by Horizzon
2. Tags_Lost to EINS
3. Tags_Will revert after reading the email
4. Lead Org_Lead Add Form
5. Last Act_SMS Sent
6. Total Time Spent on Website

- **CUTOFF**
- after analyzing ROC,sensitivity specificity tradeoff and precision recall tradeoff we decided that our cutoff should be 0.27
- That means if probability of a lead is greater than 0.27 then probably that customer get converted or convert into paying customers.
- and if the probability of a lead is less than 0.27 indicating that the customer does not convert or not buying any course .
- **our model accuracy is around 92%**
- which tells us how well we identify positives as positives and negatives as negative in our case positives means converted leads and negative means not converted leads .
- But only high accuracy don't solve our business problem .
- our main goal is to correctly identify the hot leads or converted leads .
- That means our sensitivity/recall and precision should be as high as possible .
- **our model sensitivity or recall is around 92%**
- which tells us how well we identified the actual positives as positive. In our case how well we identified the actual converted leads(hot leads) as converted leads(hot leads).
- **our model specificity is around 91%**
- which tells us how well we identified the actual negative as negative in our case negative means not converted leads.
- **our model precision is around 86%**
- which tells us that from all predicted positives how many of them are correctly identified as positive.
- **so overall our model is doing great**




Lead Score case Study End