Lead Scoring Case Study

By Sana Aafreen

Problem Statement:

- X Education is an education company sells online courses to industry professionals.
- It needs help to select the most promising leads which are most likely to convert into paying customers.
- The company needs to build a model to assign a lead such that the customers with higher lead score have a higher conversion chance and vice versa.
- The CEO has also set the target to lead conversion rate to be around 80%.

Business Goal

- X education wants to know most promising leads.
- For that they want to build a Model which identifies the hot leads.
- Deployment of the model for the future use.

Way for Solution

- Data cleaning and data manipulation: to check and handle duplicate data, to check and handle N/A values and missing values, to drop columns, if it contains large amount of missing values and not useful for the analysis, imputation of the values, if necessary, to check and handle outliers in data.
- Univariate data analysis: value count, distribution of variable etc.
- Bivariate data analysis: correlation coefficients and pattern between the variables etc.

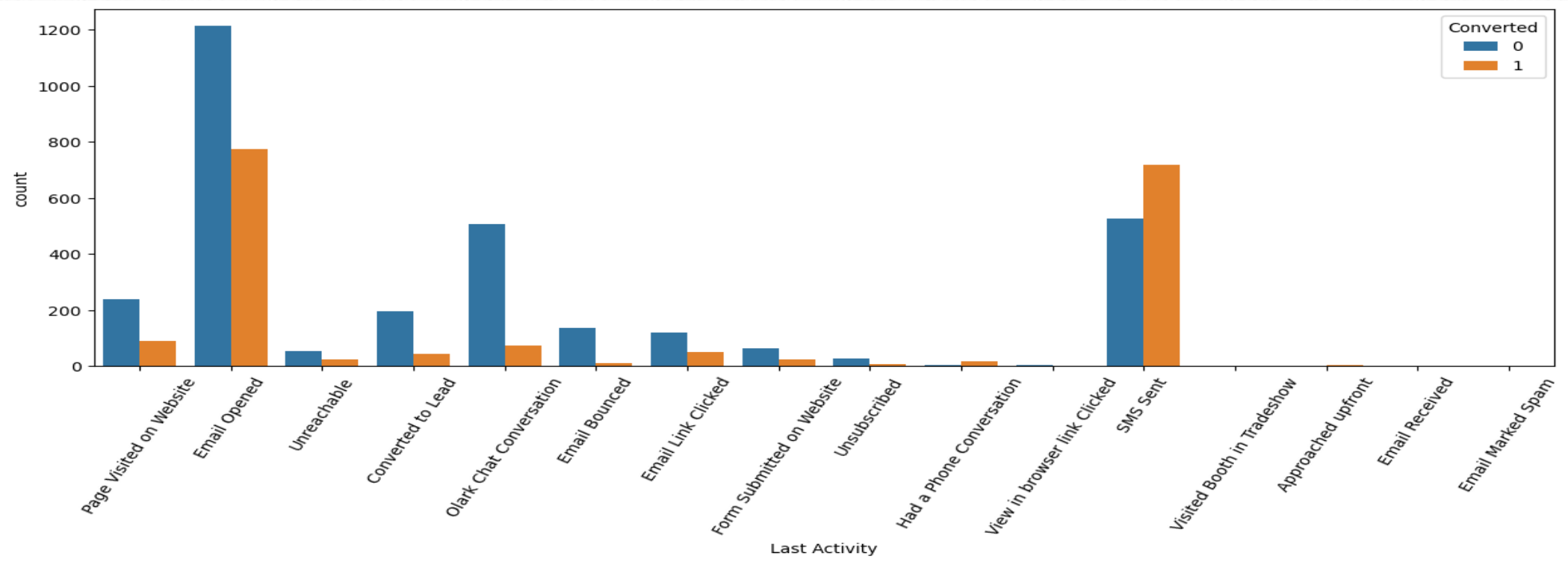
- Feature Scaling & Dummy Variables and encoding of the data.
- Classification technique: logistic regression used for the model making and prediction.

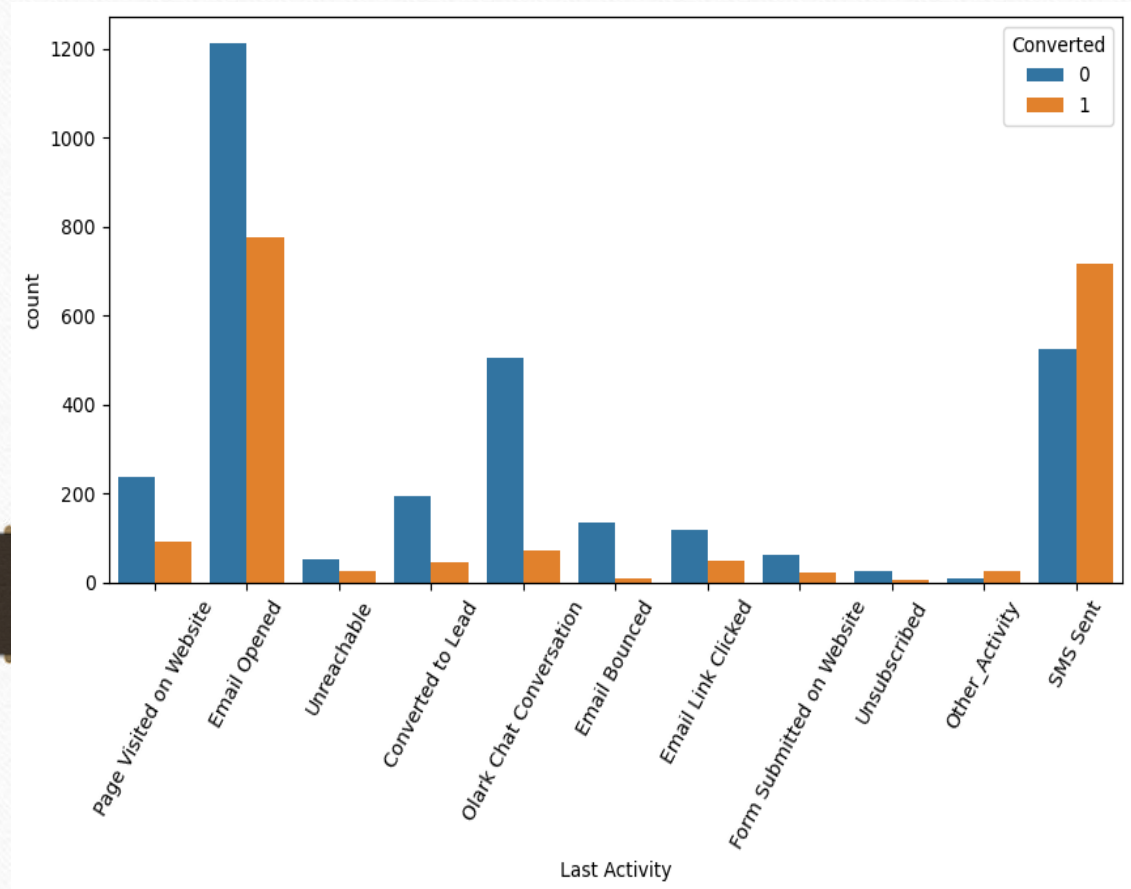
- Validation of the model.
- Model presentation.
- Conclusion

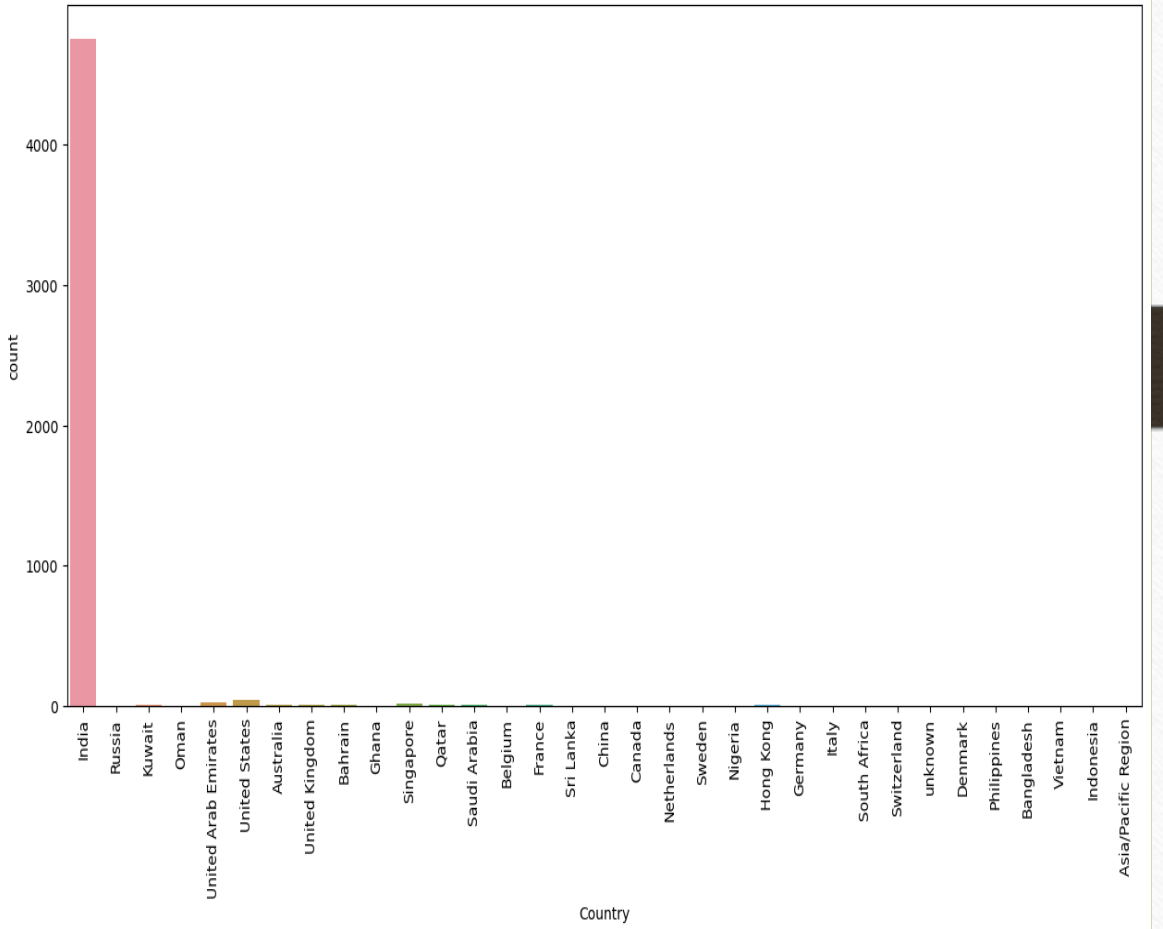
Data Manipulation

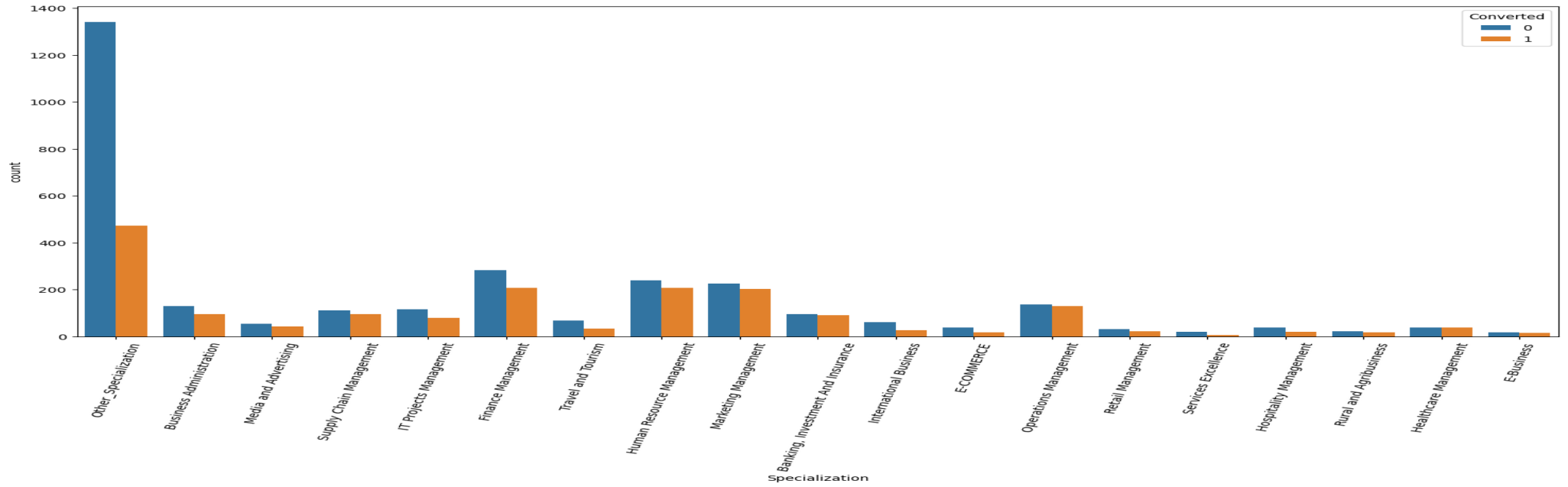
- Total Number of Rows =37, Total Number of Columns =9240.
- Based on Univariate analysis, many columns are not adding any values to the method. So it is better drop these columns.
- For categorical variables with multiple levels, create dummy features (one-hot encoded)
- Removing the “Prospect ID” which is not necessary for the analysis.

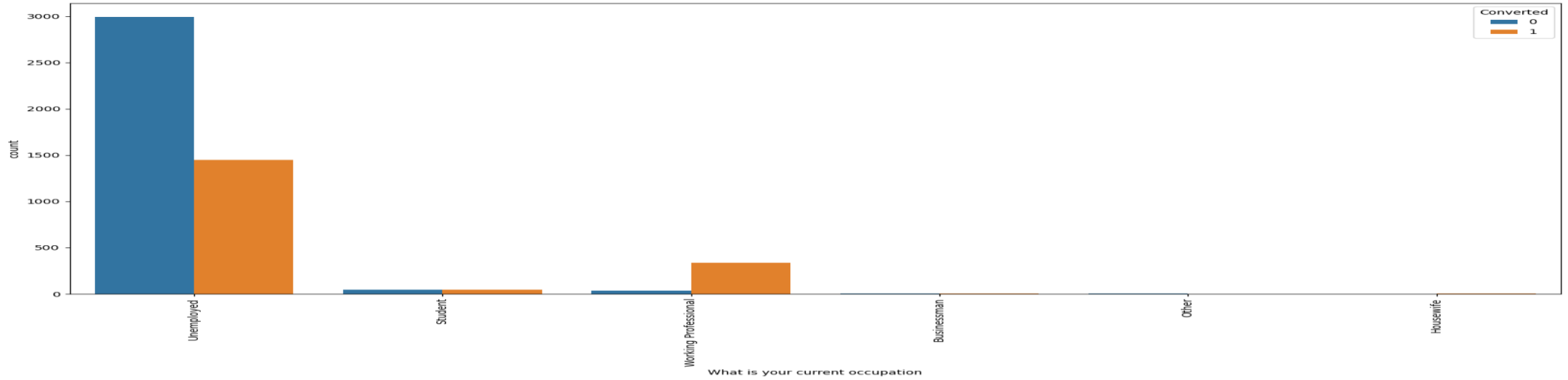
Data Analysis

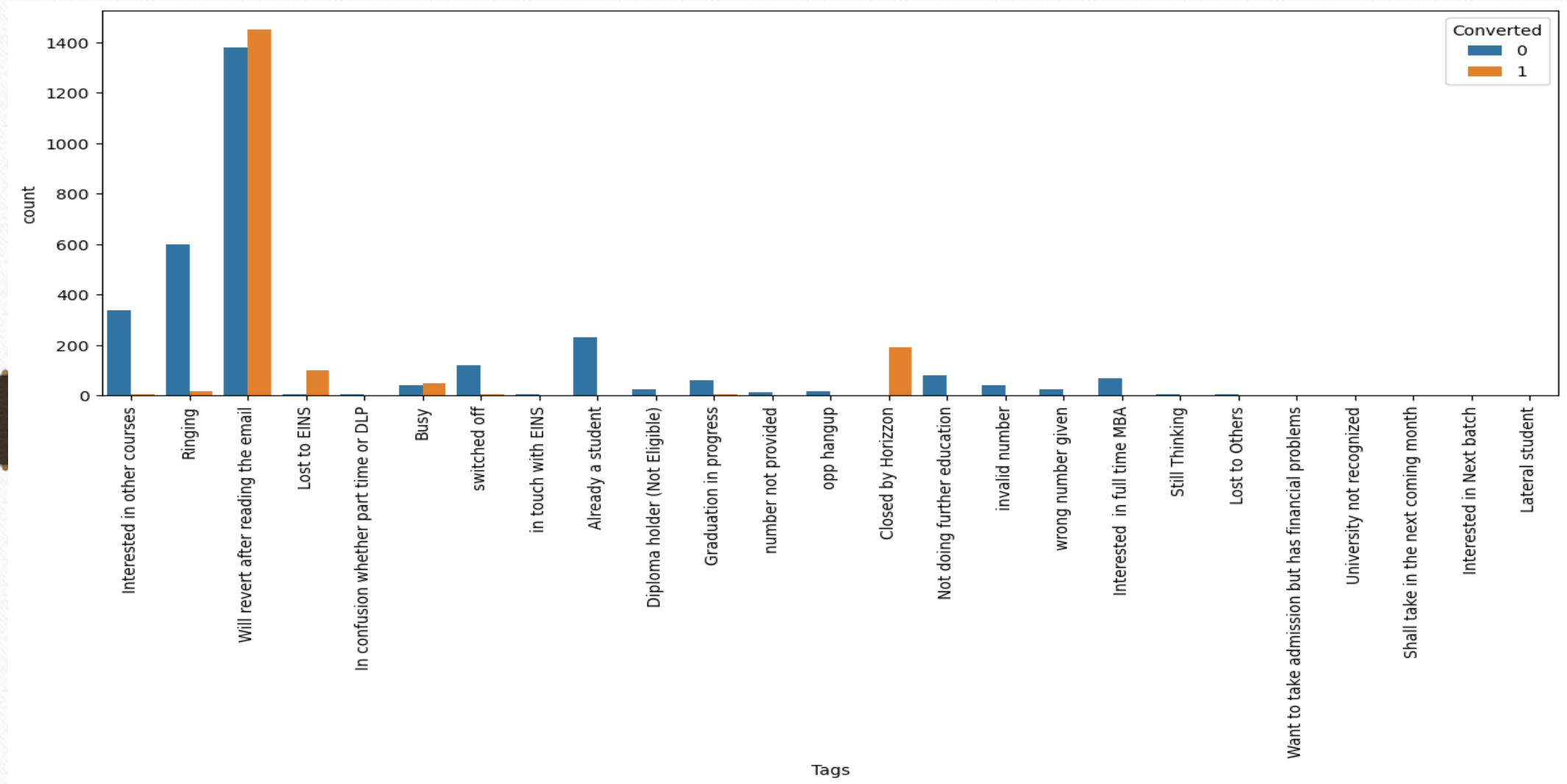


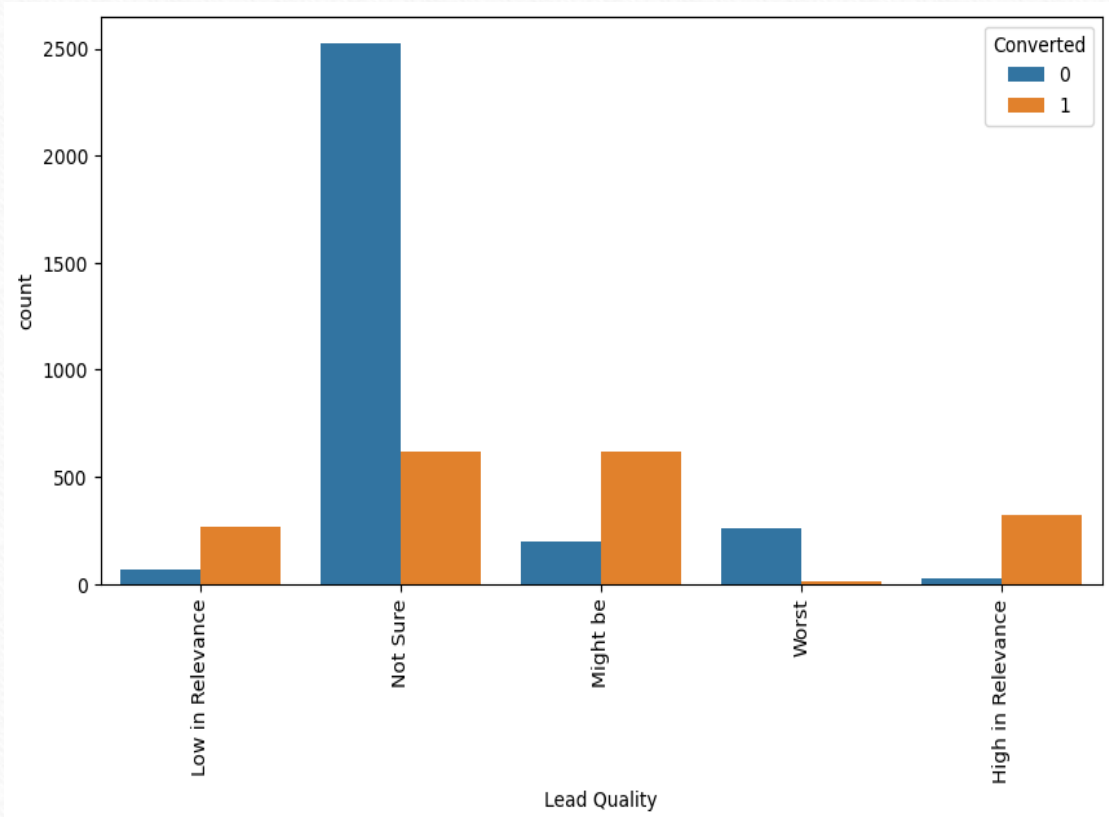

Converted
0
1
count
1200
1000
800
600
400
200
0
Page Visited on Website
Email Opened
Unreachable
Converted to Lead
Olark Chat Conversation
Email Bounced
Email Link Clicked
Form Submitted on Website
Unsubscribed
Other_Activity
SMS Sent
Last Activity

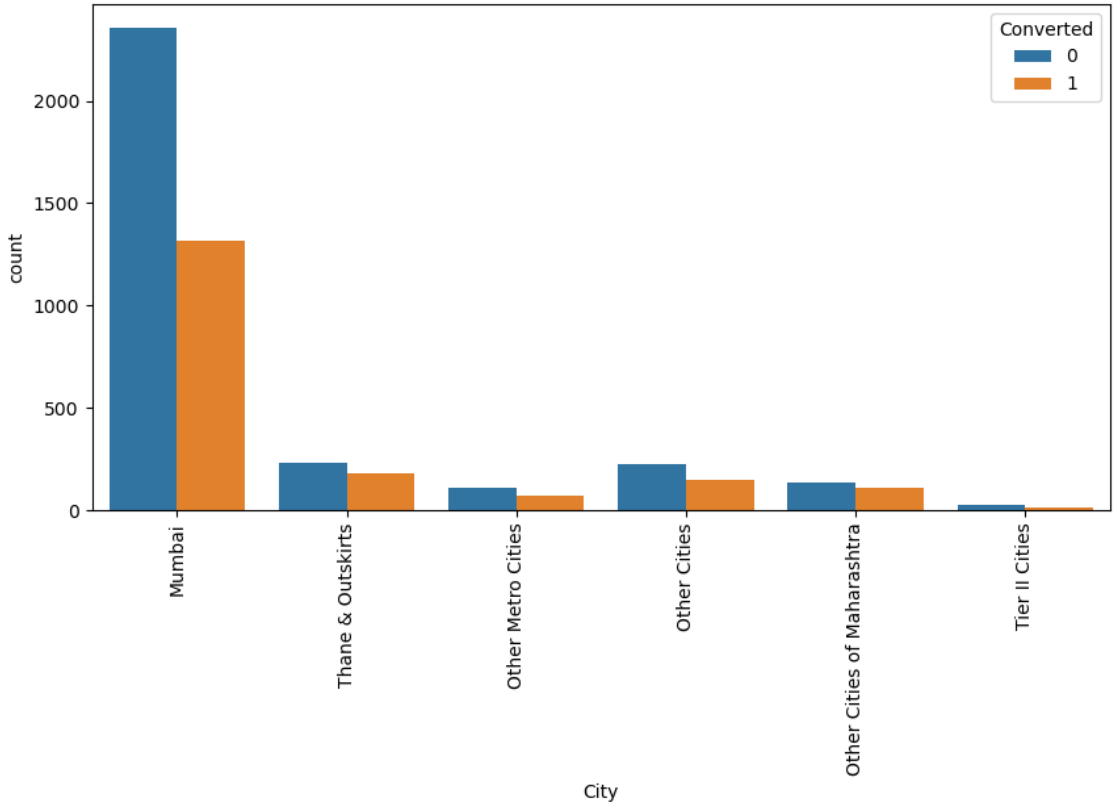

count
4000
3000
2000
1000
0
India
Russia
Kuwait
Oman
United Arab Emirates
United States
Australia
United Kingdom
Bahrain
Ghana
Singapore
Qatar
Saudi Arabia
Belgium
France
Sri Lanka
China
Canada
Netherlands
Sweden
Nigeria
Hong Kong
Germany
Italy
South Africa
Switzerland
unknown
Denmark
Philippines
Bangladesh
Vietnam
Indonesia
Asia/Pacific Region
Country

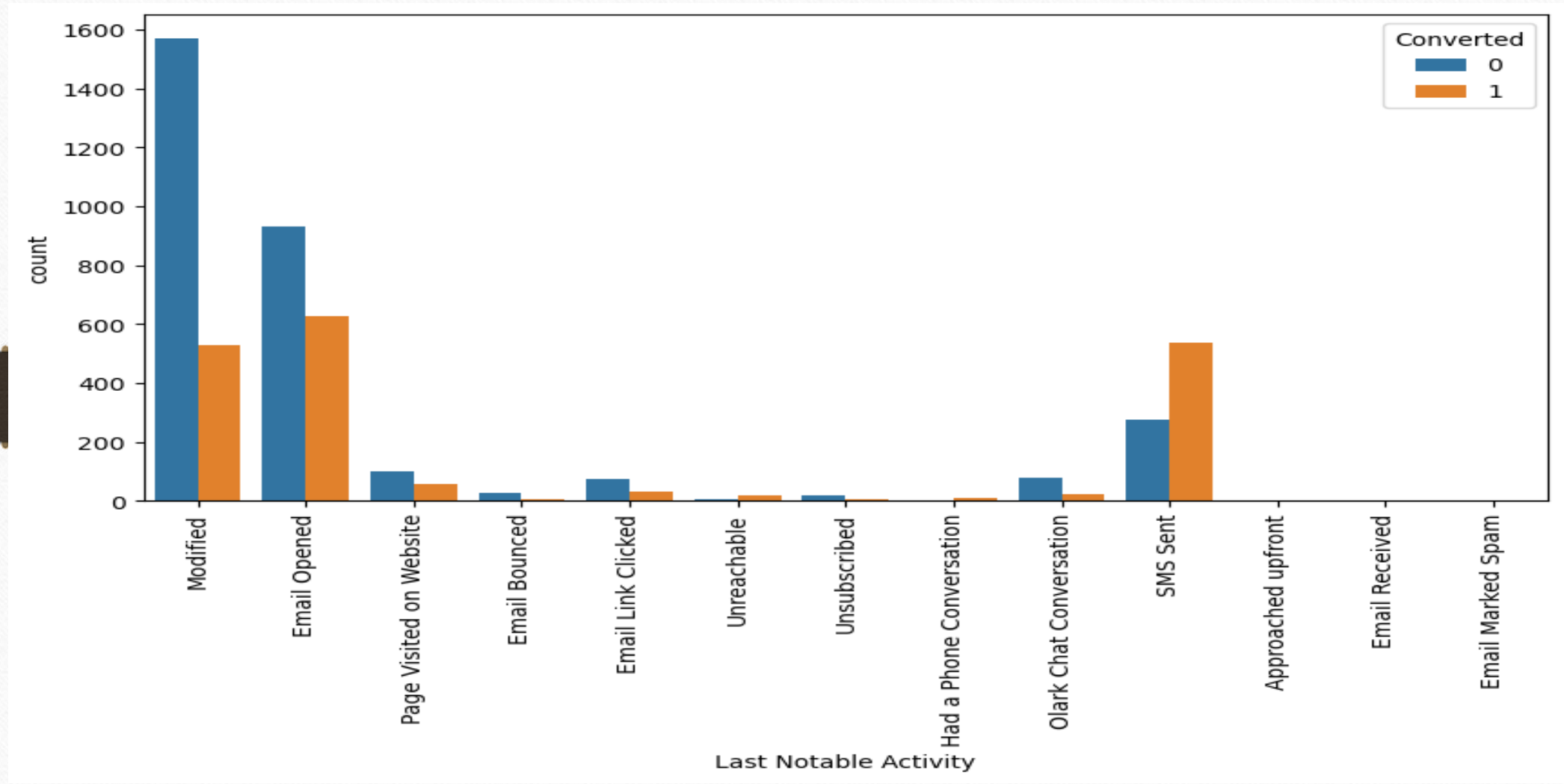

Converted
0
1
count
1400
1200
1000
800
600
400
200
0
Other_Specialization
Business Administration
Media and Advertising
Supply Chain Management
IT Projects Management
Finance Management
Travel and Tourism
Human Resource Management
Marketing Management
Banking, Investment And Insurance
International Business
E-COMMERCE
Operations Management
Retail Management
Services Excellence
Hospitality Management
Rural and Agribusiness
Healthcare Management
E-Business
Specialization


Converted
0
1
count
3000
2500
2000
1500
1000
500
0
Unemployed
Student
Working Professional
Businessman
Other
Housewife
What is your current occupation


Converted
0
1
count
1400
1200
1000
800
600
400
200
0
Interested in other courses
Ringing
Will revert after reading the email
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
number not provided
opp hangup
Closed by Horizzon
Not doing further education
invalid number
wrong number given
Interested in full time MBA
Still Thinking
Lost to Others
Want to take admission but has financial problems
University not recognized
Shall take in the next coming month
Interested in Next batch
Lateral student
Tags

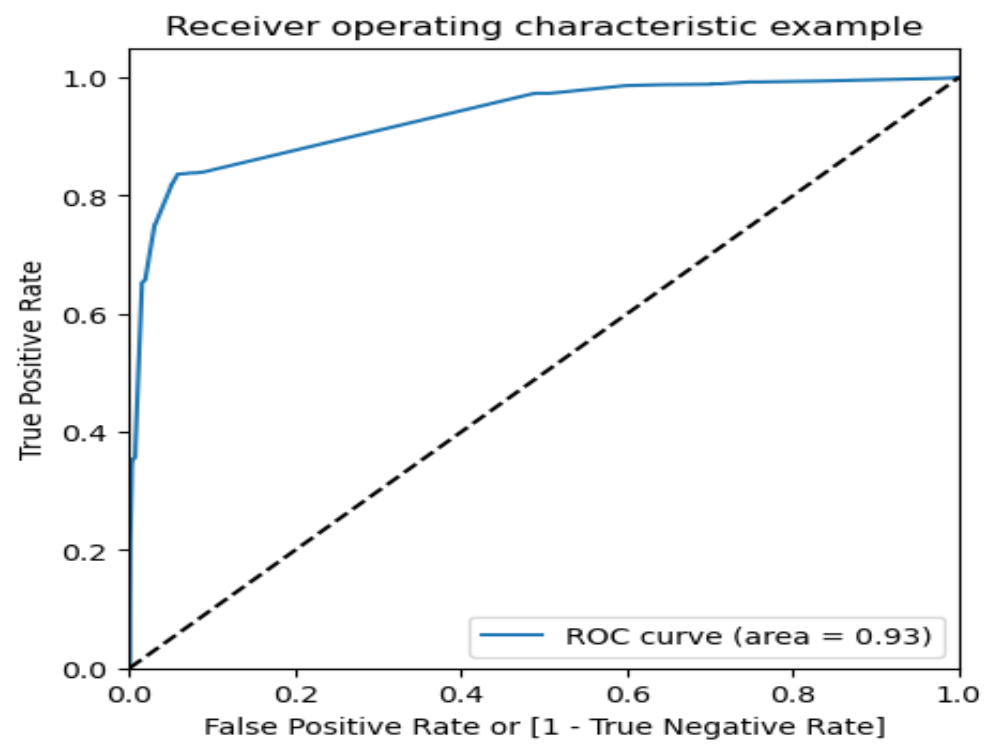

Converted
0
1
count
2500
2000
1500
1000
500
0
Low in Relevance
Not Sure
Might be
Worst
High in Relevance
Lead Quality


Converted
0
1
count
2000
1500
1000
500
0
Mumbai
Thane & Outskirts
Other Metro Cities
Other Cities
Other Cities of Maharashtra
Tier II Cities
City


Converted
0
1
count
1600
1400
1200
1000
800
600
400
200
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
Email Received
Email Marked Spam
Last Notable Activity

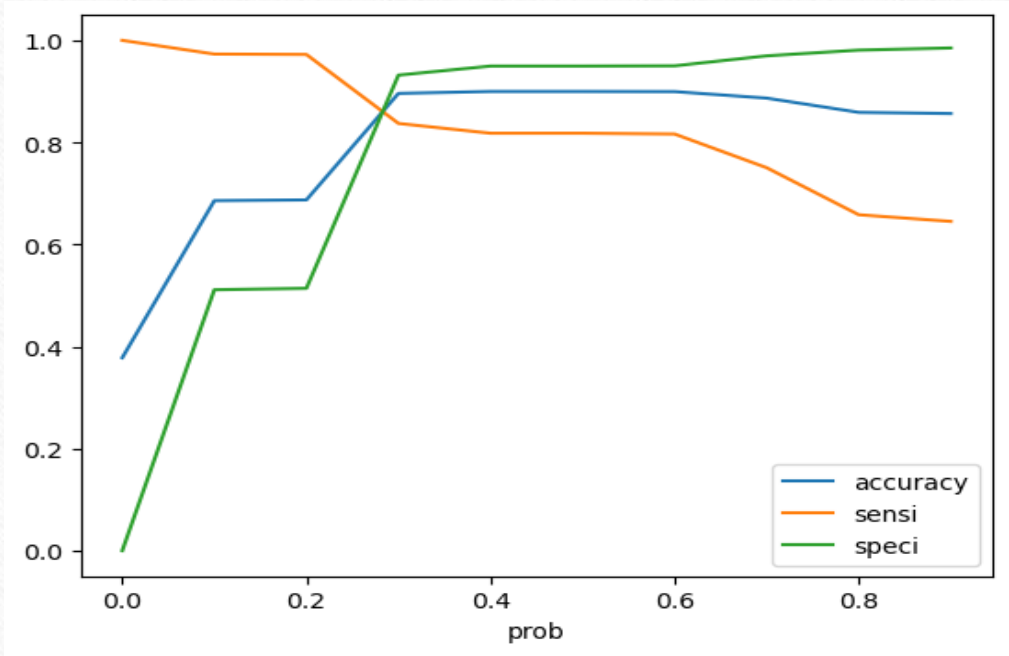
Receiver Operating Characteristics

- It shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity).
- The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.
- The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.



Optimal Cut off Point

- Finding Optimal Cut off Point
- Optimal cut off probability is that probability where we get balanced sensitivity and specificity.
- From the second graph it is visible that the optimal cut off is at 0.3.



Conclusion

According to the research leads can be converted to potential buyer when:

- When their current occupation is as a working professional.
- When the lead origin is Lead add format.
- When the lead source was: a. Google b. Direct traffic
- When the last activity was: a. SMS b. Olark chat conversation
- When the total number of visits in website is more.