Lead Scoring Case Study

By:
Shrey Khurana
&
Daman

Problem Statement

- X Education is an Education Company that sells online courses to industry professionals. Professionals who are interested in the courses land on their website and browse for courses. The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead.

- X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need 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. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Exploratory Data Analysis (EDA)

Dividing the Variables for EDA:

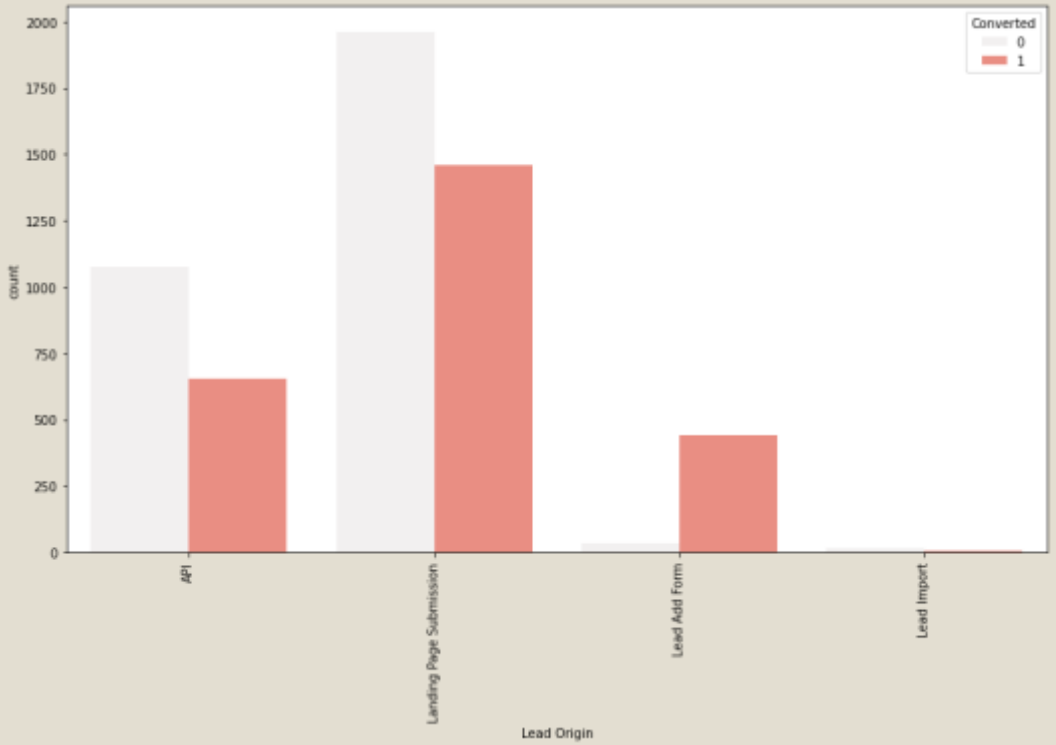
- ✓ **Categorical Columns:** Lead Origin,Lead Source,Do Not Email,Do Not Call,Last Activity,Specialization,What is your current occupation,What matters most to you in choosing a course,Search,X Education Forums,Newspaper,Digital Advertisement,Through Recommendations,Receive More Updates About Our Courses,City,A free copy of Mastering The Interview,Last Notable Activity

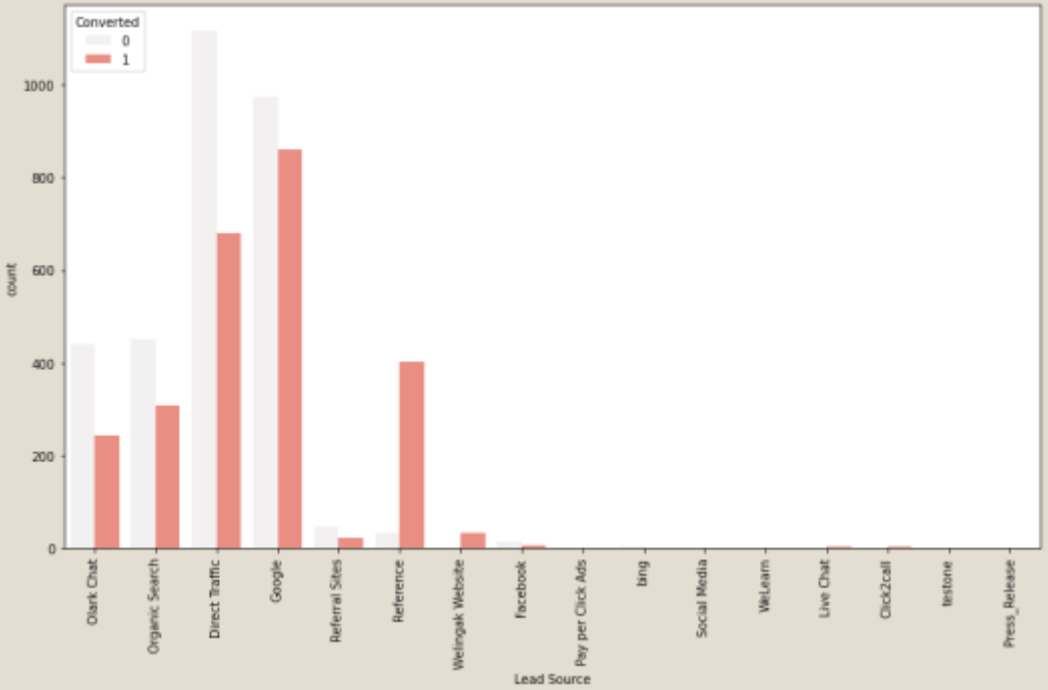
- ✓ **Continuous Columns:** Converted,TotalVisits,Total Time Spent on Website,Page Views Per Visit

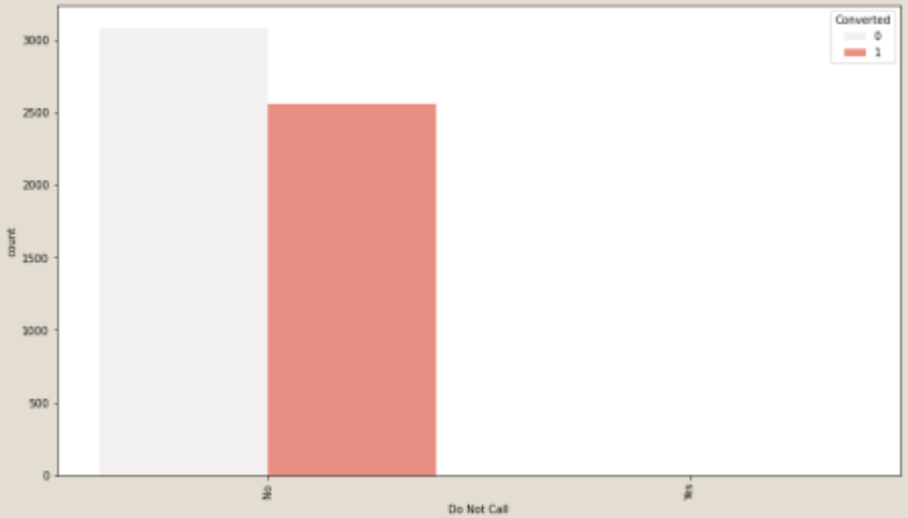
Univariate Analysis of Categorical Columns:

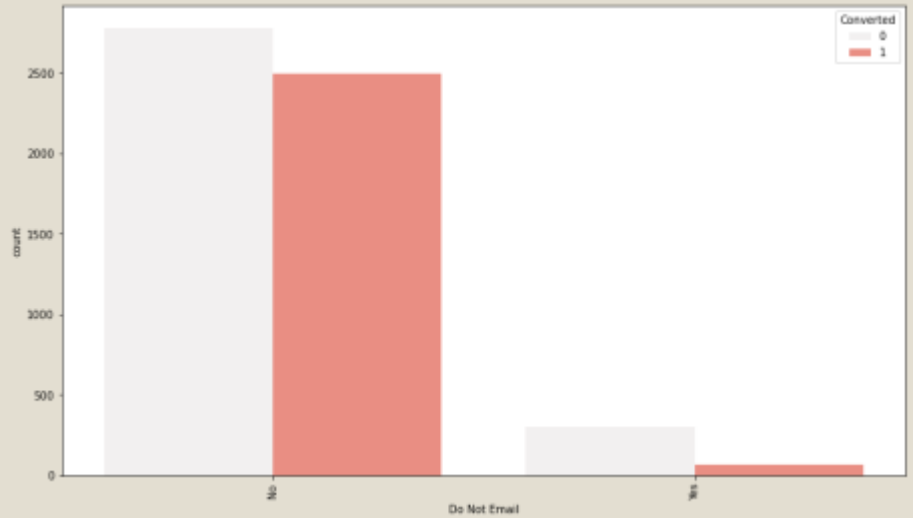
- ✓ Lead Origin is highest for '**Landing Page Submission**'.
- ✓ Majority of the leads are from '**Google**' followed by 'Direct Traffic'.
- ✓ Majority of users have **not** opted for Email.
- ✓ Majority of users have **not** opted for Call.
- ✓ Most of the users Last Activity is that they have opened Email and then followed by SMS.
- ✓ Out of all the specializations offered, **Management Course** are in great demand.
- ✓ There are large number of Unemployed.
- ✓ '**Better Career Prospects**' plays a major role while choosing the course.
- ✓ Almost no one wants to receive the updates about courses.
- ✓ Majority of Users are from **Mumbai**, followed by users have not mentioned there city.
- ✓ Most of the users **don't want** a free copy of Mastering Interview.
- ✓ The highest number of Last Notable activities are - **Modified, Email Opened, SMS Sent**.

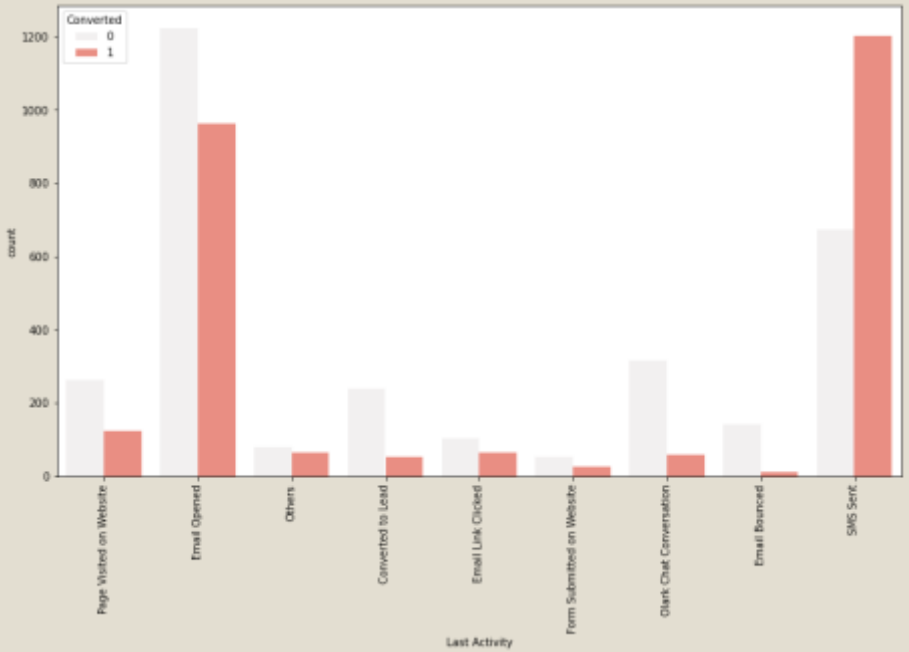
Bivariate Analysis:

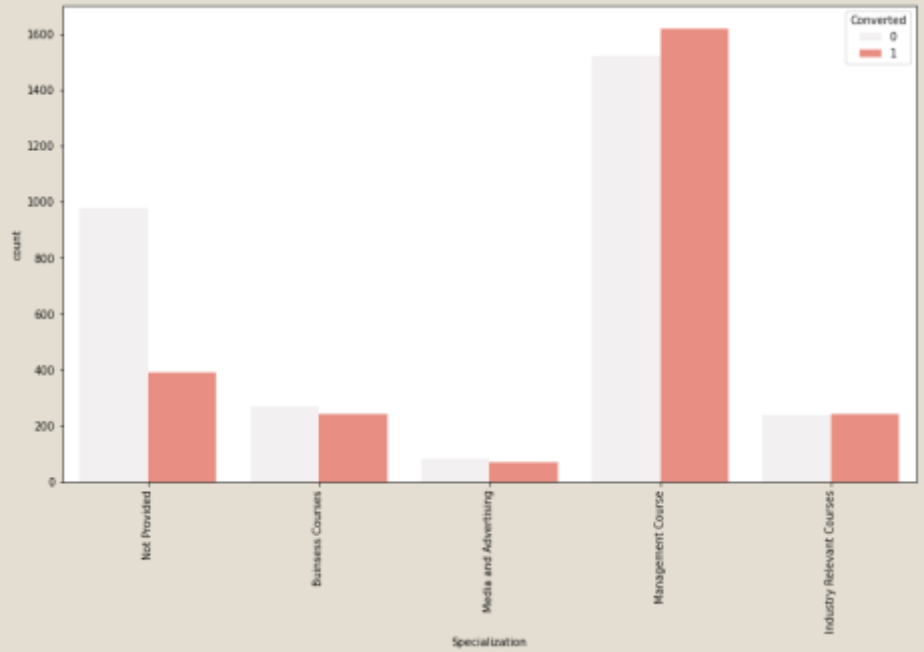


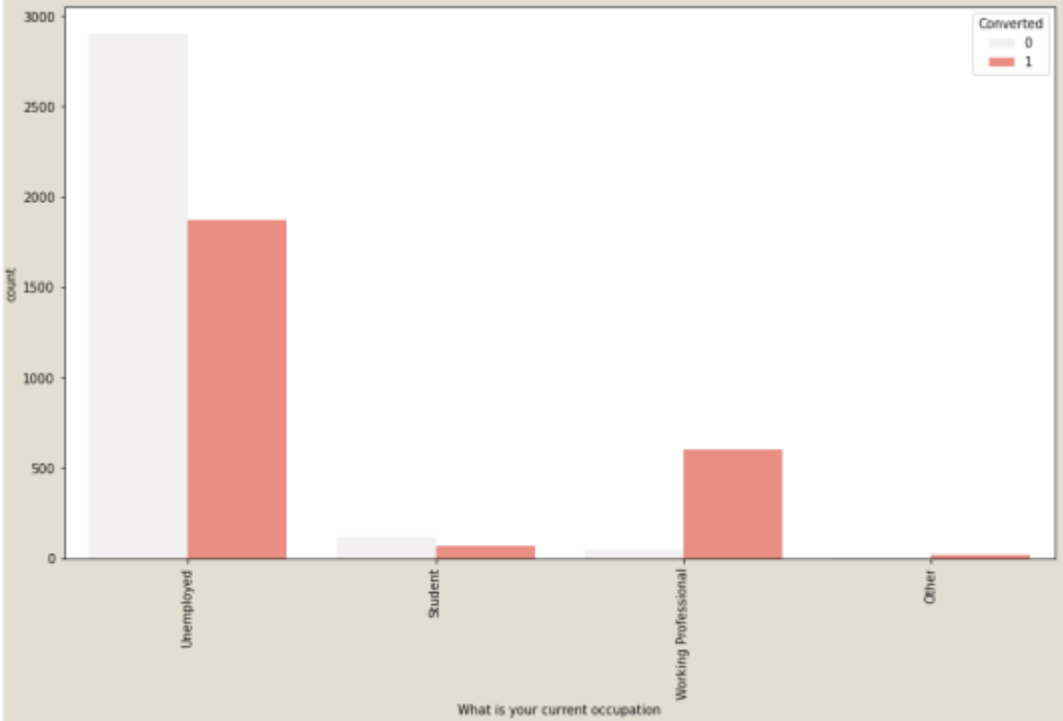


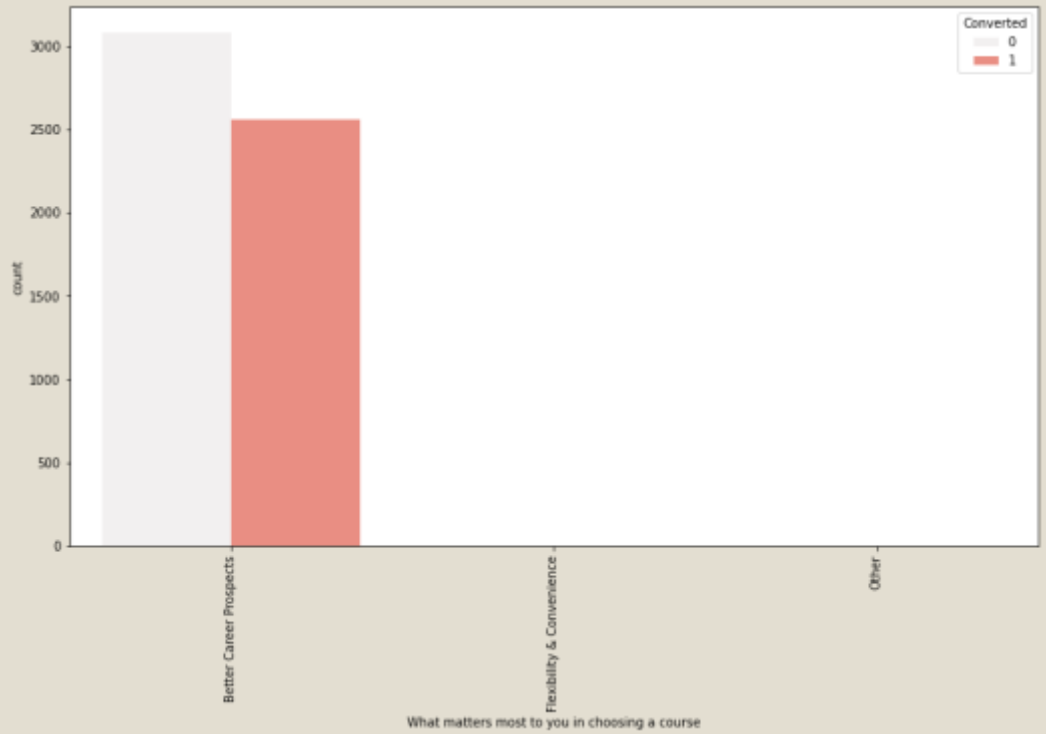

Converted
0
1
count
3000
2500
2000
1500
1000
500
0
No
Yes
Do Not Call

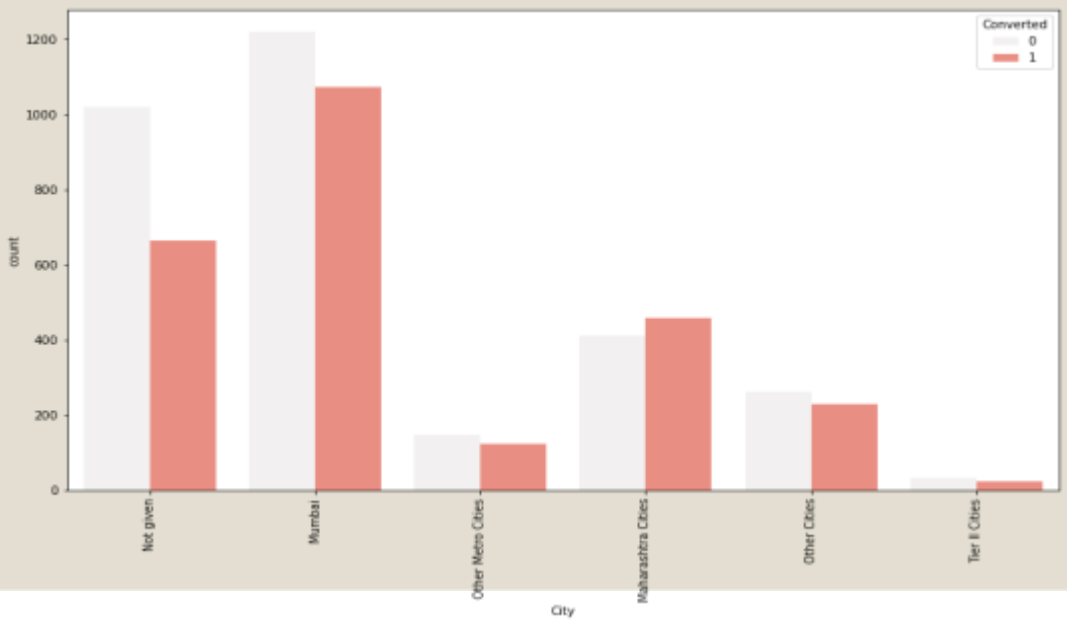

Converted
0
1
count
2500
2000
1500
1000
500
0
No
Yes
Do Not Email

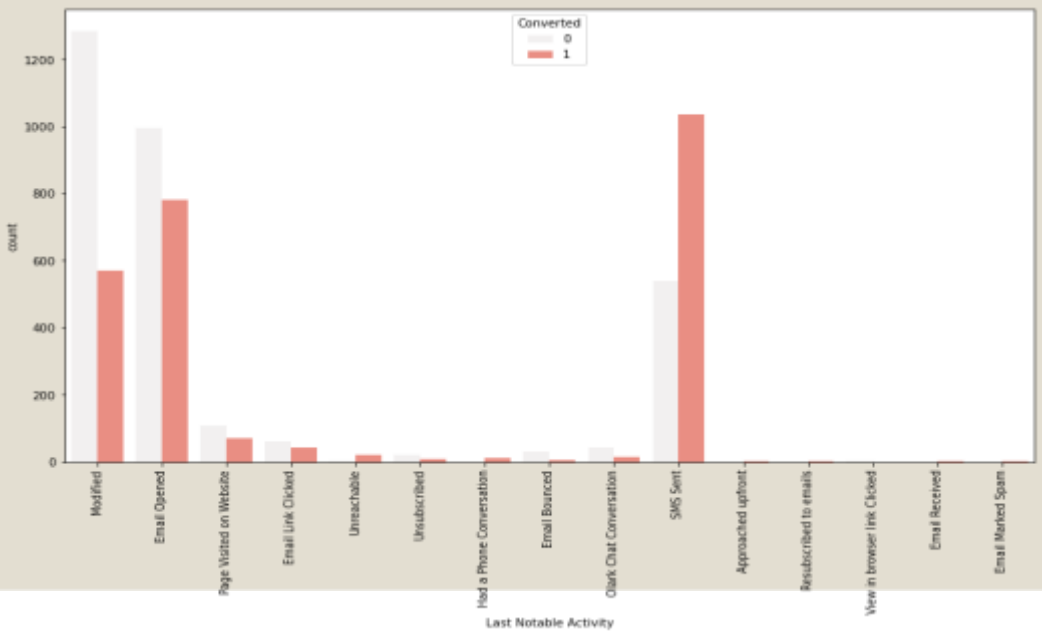

Converted
0
1
count
1200
1000
800
600
400
200
0
Page Visited on Website
Email Opened
Others
Converted to Lead
Email Link Clicked
Form Submitted on Website
Olark Chat Conversation
Email Bounced
SMS Sent
Last Activity


Converted
0
1
count
1600
1400
1200
1000
800
600
400
200
0
Not Provided
Business Courses
Media and Advertising
Management Course
Industry Relevant Courses
Specialization


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


Converted
0
1
count
3000
2500
2000
1500
1000
500
0
Better Career Prospects
Flexibility & Convenience
Other
What matters most to you in choosing a course


Converted
0
1
count
1200
1000
800
600
400
200
0
Not given
Mumbai
Other Metro Cities
Maharashtra Cities
Other Cities
Tier II Cities
City


Converted
0
1
count
1200
1000
800
600
400
200
0
Modified
Email Opened
Page Visited on Website
Email Link Clicked
Unreachable
Unsubscribed
Had a Phone Conversation
Email Bounced
Olark Chat Conversation
SMS Sent
Approached upfront
Resubscribed to emails
View in browser link Clicked
Email Received
Email Marked Spam
Last Notable Activity

Inference from Above Plots:

- ✓ **Landing on the Page Submission** gets the most converted.
- ✓ Leads from **Direct Traffic** converts the most and then followed by **Google.**
- ✓ Users who '**Opened Mail**' gets converted the most.
- ✓ Users who '**Sent SMS**' gets converted the most.
- ✓ **Unemployed** and **Working Professionals** gets converted.
- ✓ People who want Better career prospects will choose a course.
- ✓ People with No cities Provided and living in Mumbai or other Maharashtra Cities will take the course.
- ✓ Users who have said **no to Newspaper** and **Digital Advertisement** will get converted.
- ✓ Users who have said **No to Email** and **Calls** will get converted.

Data Preparation

Steps:

1. Dropping the Unwanted Columns.

1. Converting the columns to Binary form.

2. Creating Dummy columns.

3. Splitting the data to test and Train. Train data – 70 % and test data – 30 %.

4. Using Recursive Feature Elimination (RFE) we select 15 variables (i.e. 'Lead_Origin_Landing Page Submission', 'Lead_Origin_Lead Add Form','Lead_Source_Olark Chat', 'Lead_Source_Reference','Lead_Source_Welingak Website', 'Last_Activity_Email Bounced', 'Last_Activity_Email Opened', 'Last_Activity_Others','Last_Activity_SMS Sent', 'Specialization_Not Provided','Occu_Working Professional','Last_Notable_Activity_Had a Phone Conversation', 'Last_Notable_Activity_Unreachable', 'City_Not given', 'Total Time Spent on Website')

Model Building

- We started calculating and checking the p-value and VIF of different columns.
- We looked for the columns whose p-value is greater than 0.05 and dropping these columns.
- We built total 5 models to freeze this model building process. The 5 model had P-value and VIF value in control for all the columns.
- Then we made confusion matrix to check the accuracy and sensitivity of the model.
- We also plotted ROC curve, the area of the curve was 0.86.
- Through graph, we chose the optimal point for cut off probability as 0.4.

Comparing the Train and Test Set

Outputs from the Train Set:

Accuracy – 78.94%
Sensitivity – 79.16%
Specificity – 78.76%

Outputs from the Test Set:

Accuracy – 80.22%
Sensitivity – 81.53%
Specificity – 79.16%

Summary

In Order to have more people enroll the courses we should keep leads informed about the new courses either through mails or SMS. We can also introduce the plan of referral program where already customers can aske there friends or family members to enroll for the course and earn cash or gift coupons.We can also the Service Desk team to call the leads asking them to join the courses.

Thank You