LEAD_SCORE_ CASE_STUDY

LOGISTIC_REGRESSION

Problem Statement:

- An X Education need help to select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires us 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%.

Goals and Objectives

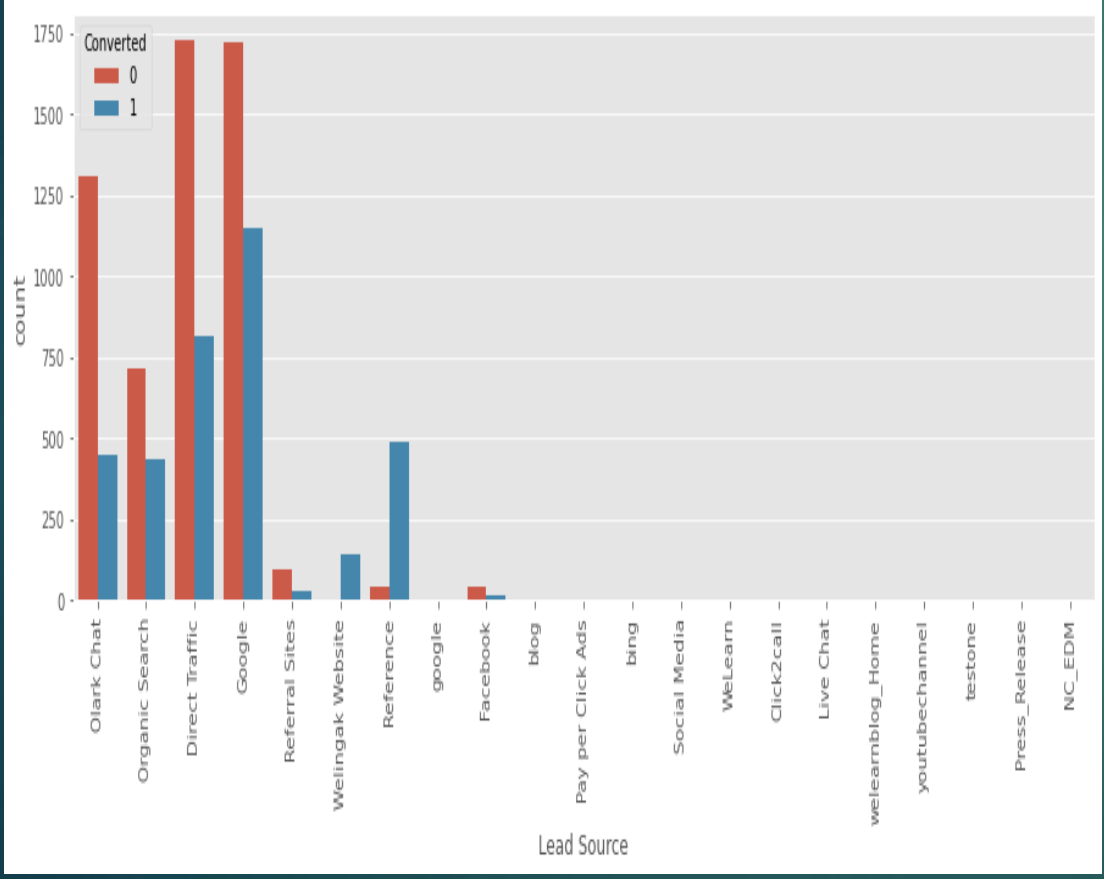
- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

STEPS PERFORMED FOR ANALYSIS:

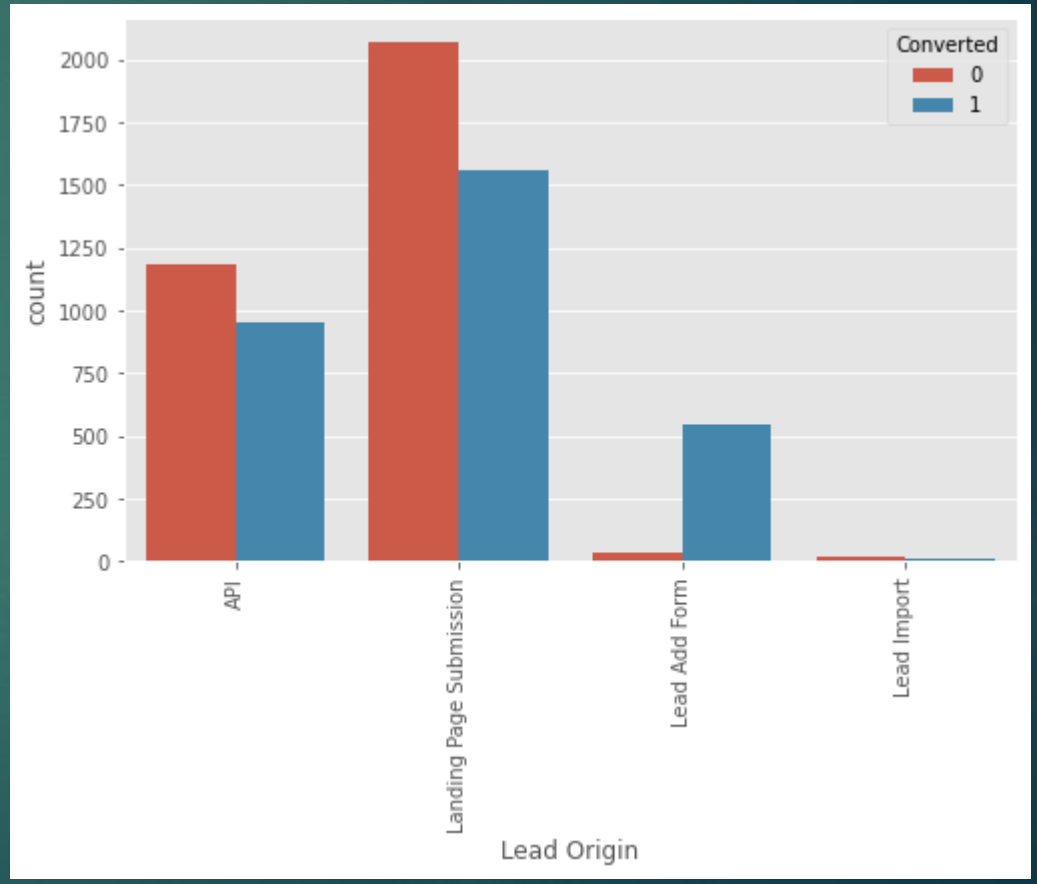
- Importing the required libraries and the dataset
- Inspecting the Data Frame
- Checking for missing values
- Analyzing columns individually and handling missing values
- Outlier treatment
- Dropping redundant columns
- Converting some binary variables (Yes/No) to 0/1
- Grouping column features
- Dummy creation
- Train - Test Split
- Model Building
- Plotting the ROC Curve
- Finding Optimal Cut-Off Points
- Precision and Recall
- Making Predictions on the Test Set

EDA

LEAD SCORE

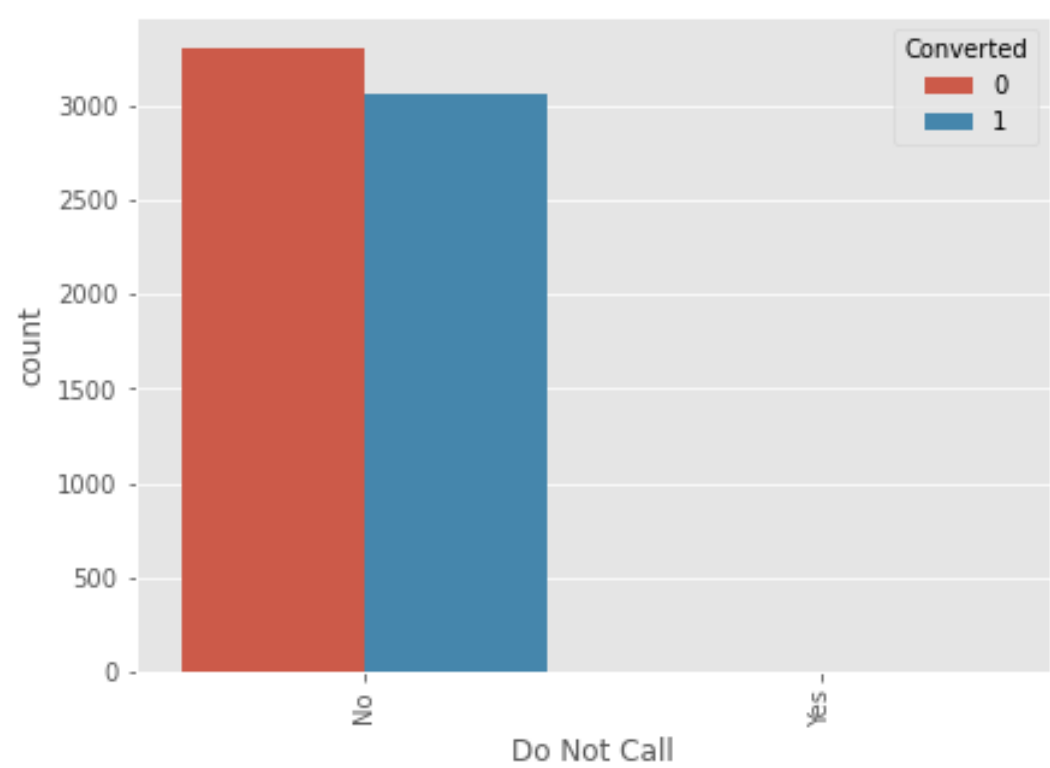


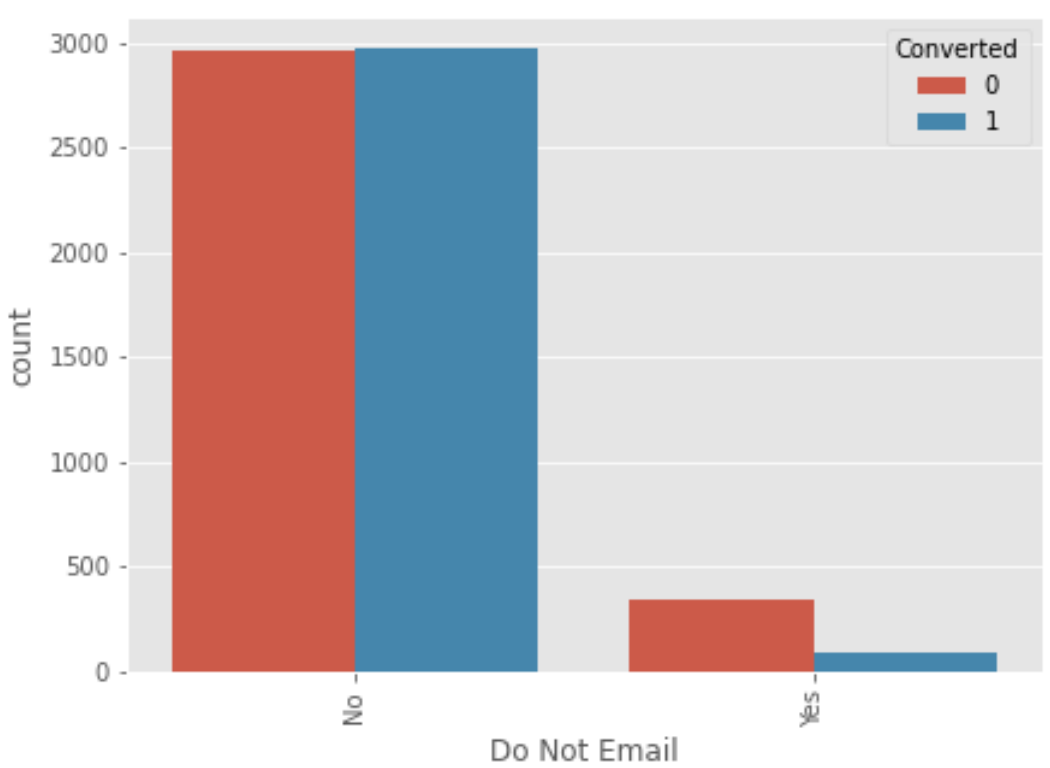
LEAD ORIGIN



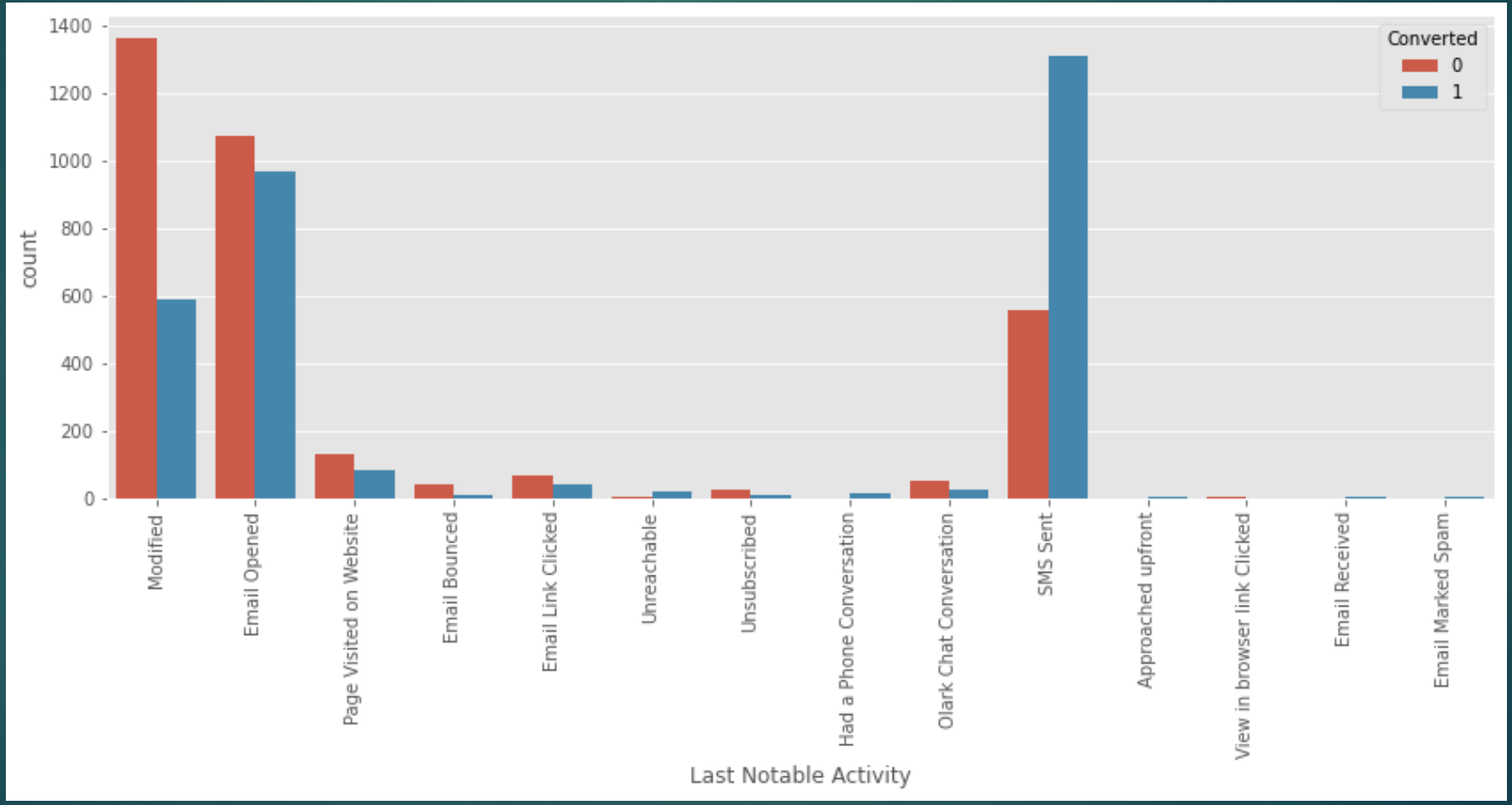
DO NOT CALL

DO NOT EMAIL

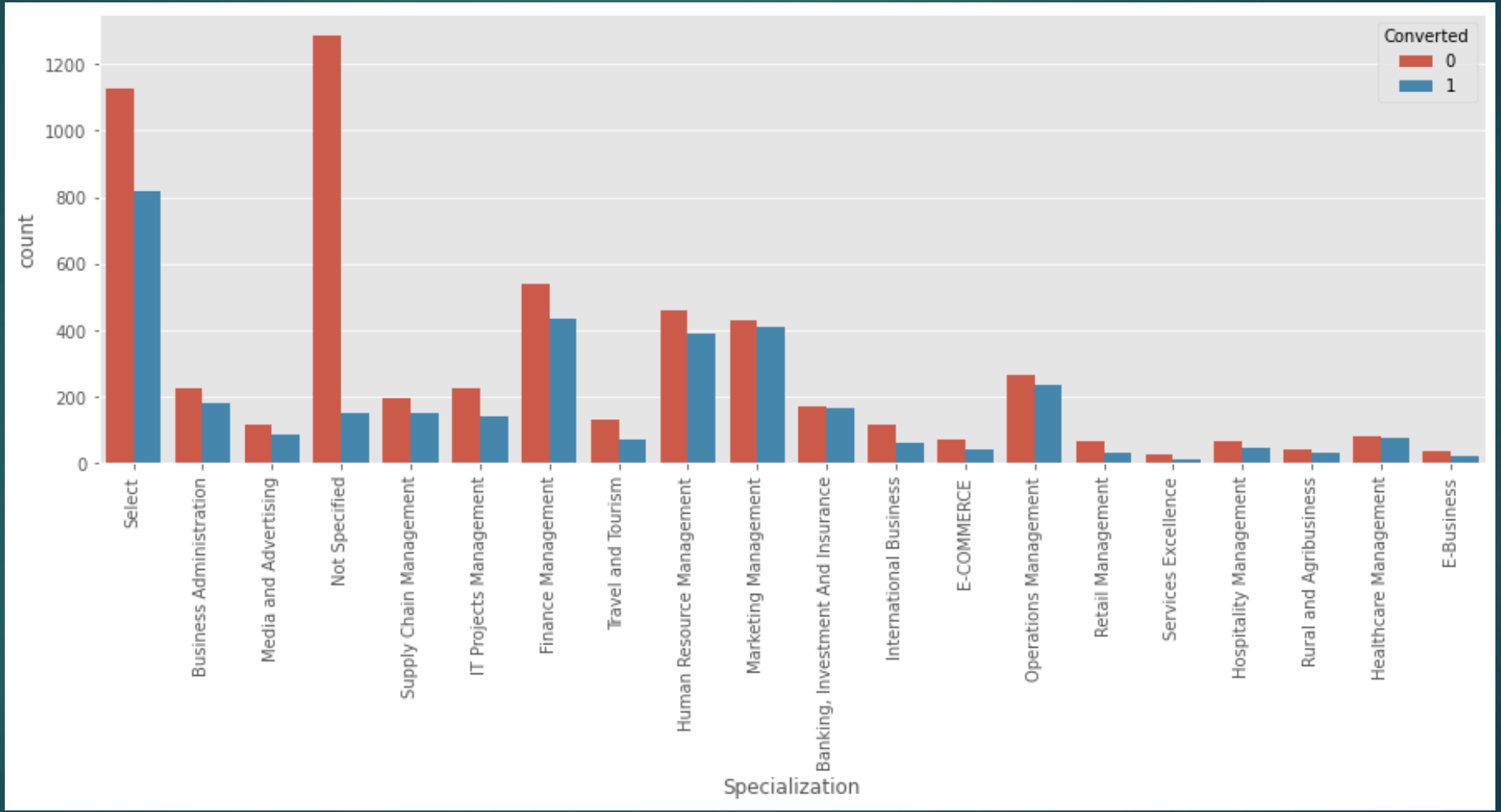




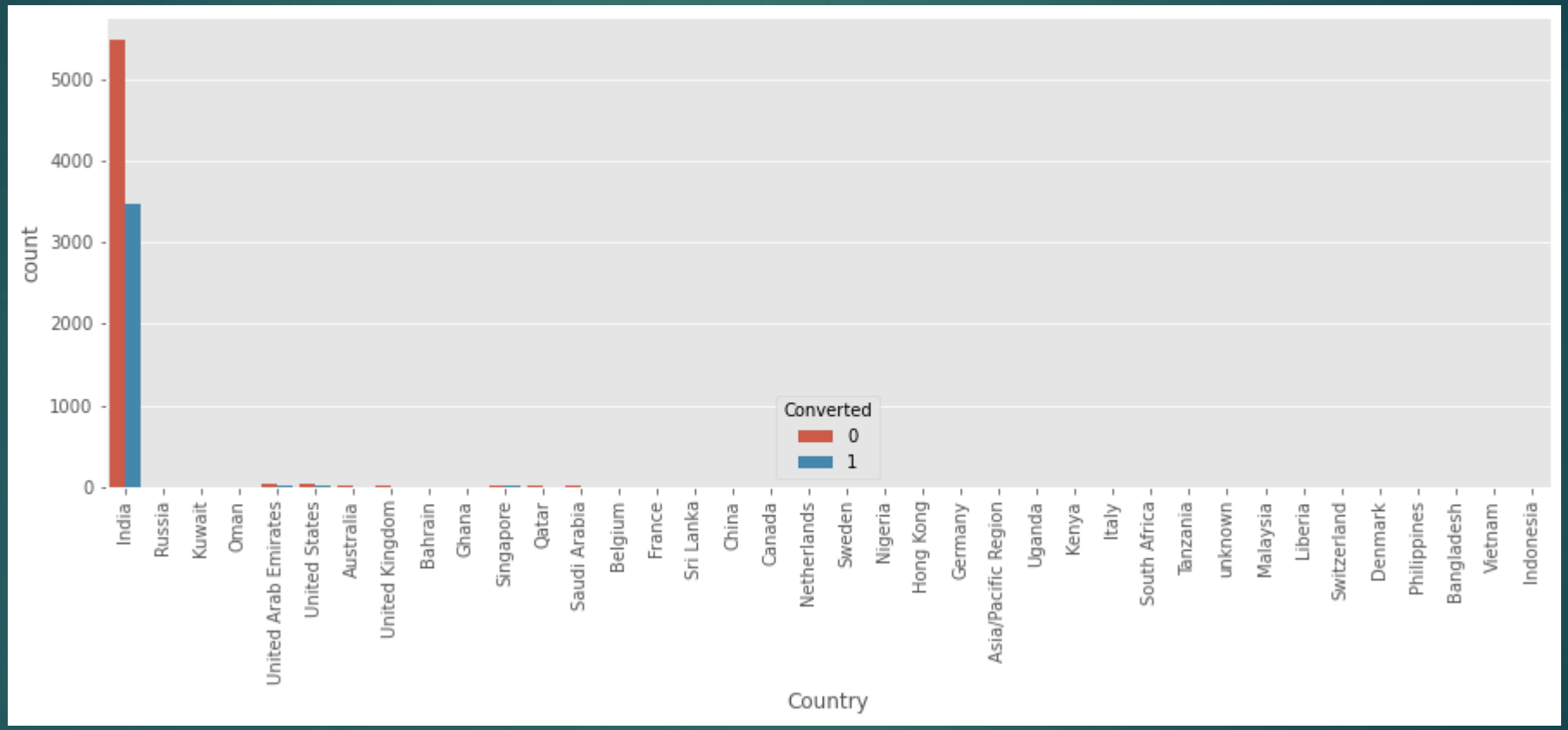
LAST ACTIVITY



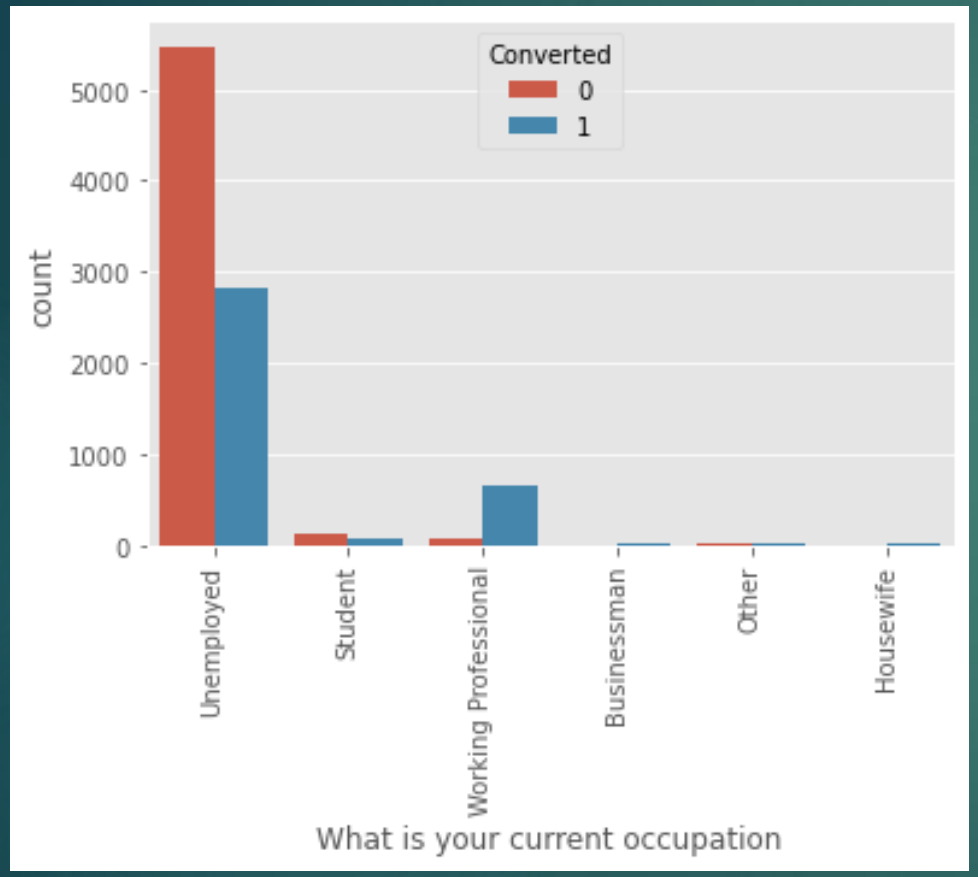
SPECIALIZATION



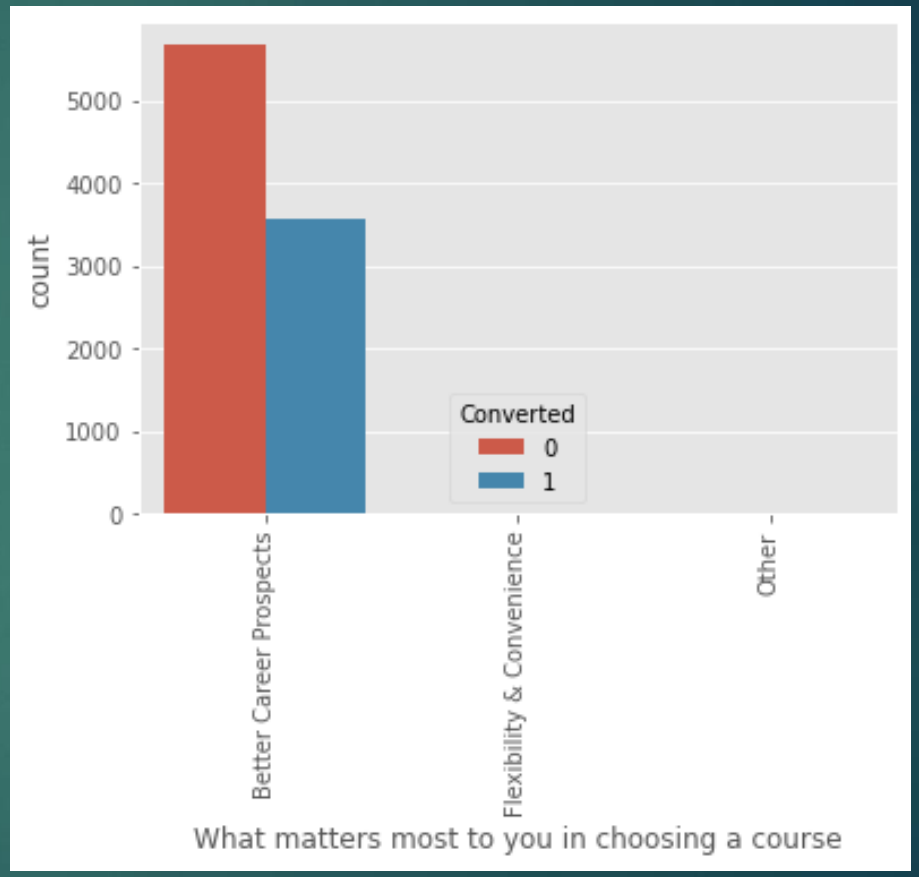
COUNTRY


count
5000
4000
3000
2000
1000
0
Converted
0
1
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
Asia/Pacific Region
Uganda
Kenya
Italy
South Africa
Tanzania
unknown
Malaysia
Liberia
Switzerland
Denmark
Philippines
Bangladesh
Vietnam
Indonesia
Country

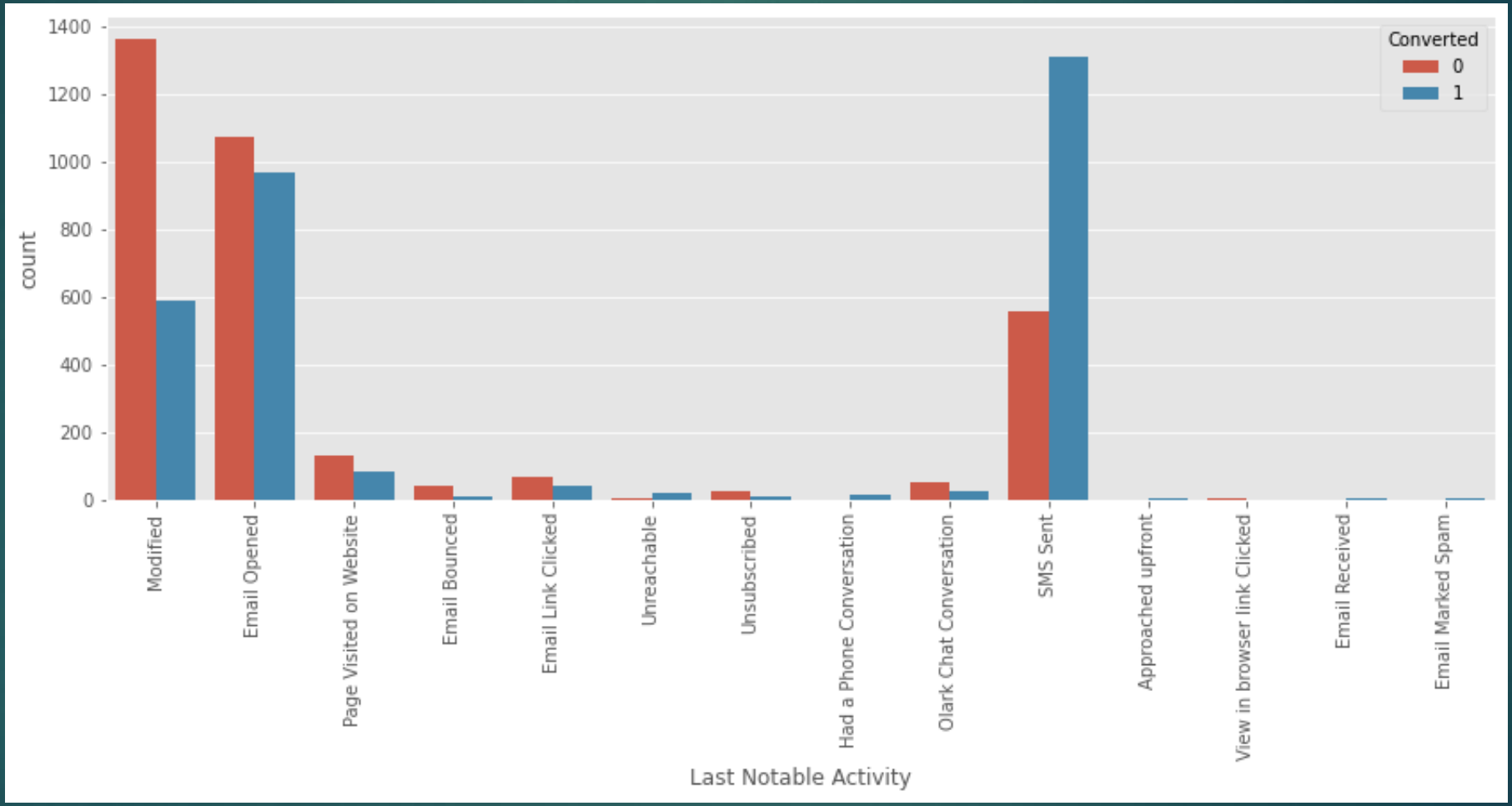
CURRENT OCCUPATION



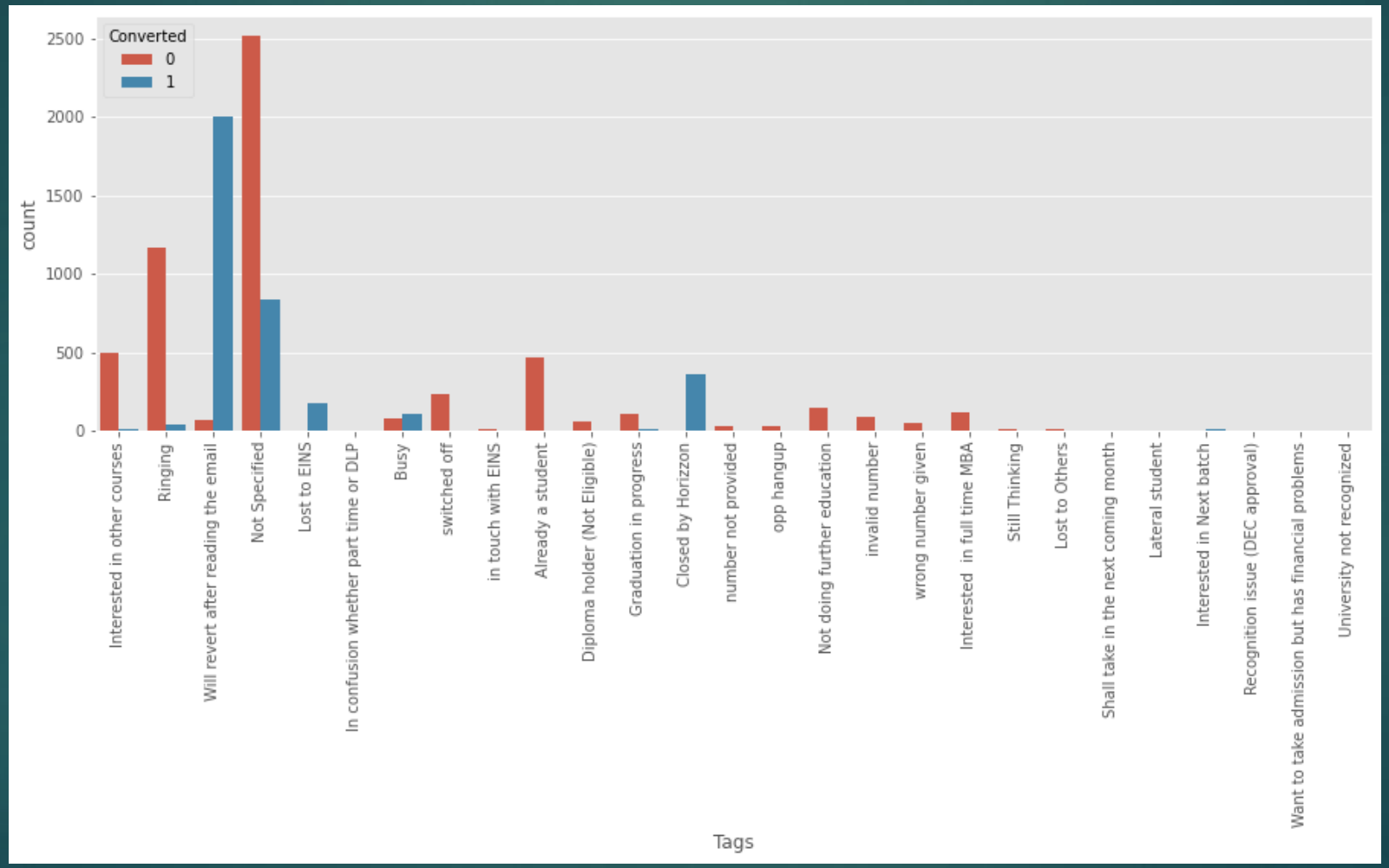
WHAT MATTERS THE MOST IN CHOOSING A COURSE ?



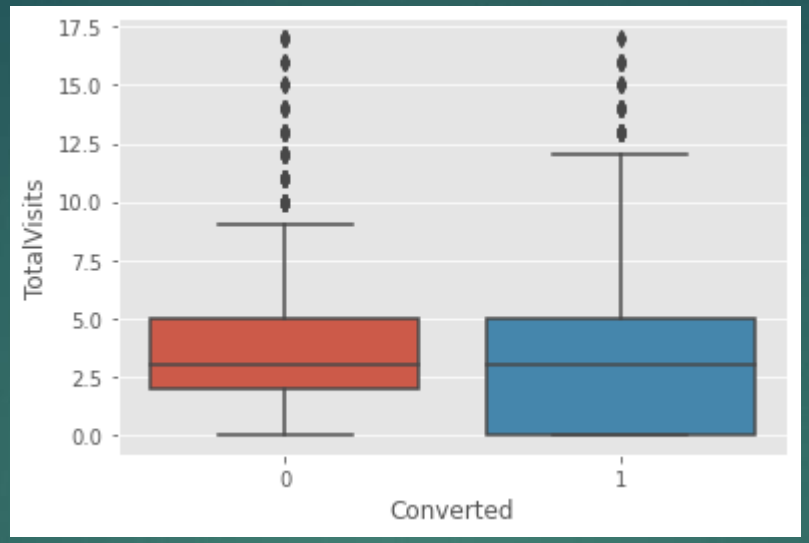
LAST NOTABLE ACTIVITY



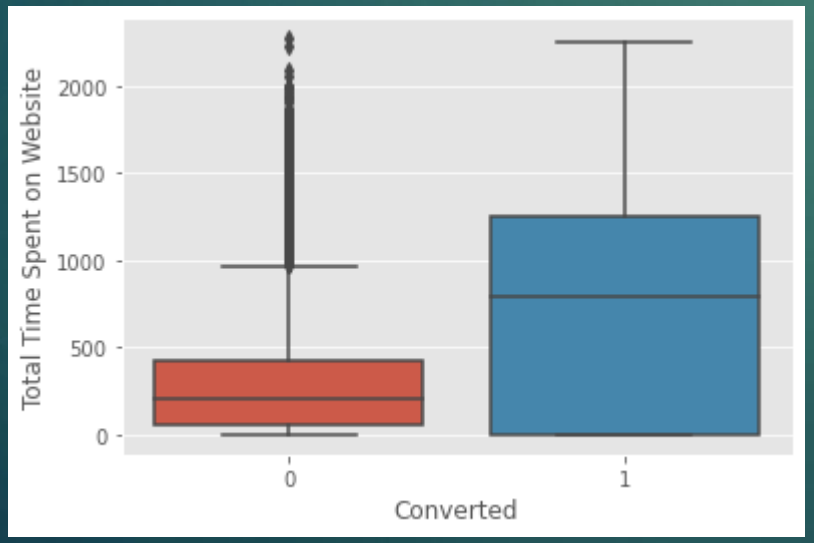
TAGS


Converted
0
1
count
2500
2000
1500
1000
500
0
Interested in other courses
Ringing
Will revert after reading the email
Not Specified
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)
Want to take admission but has financial problems
University not recognized
Tags

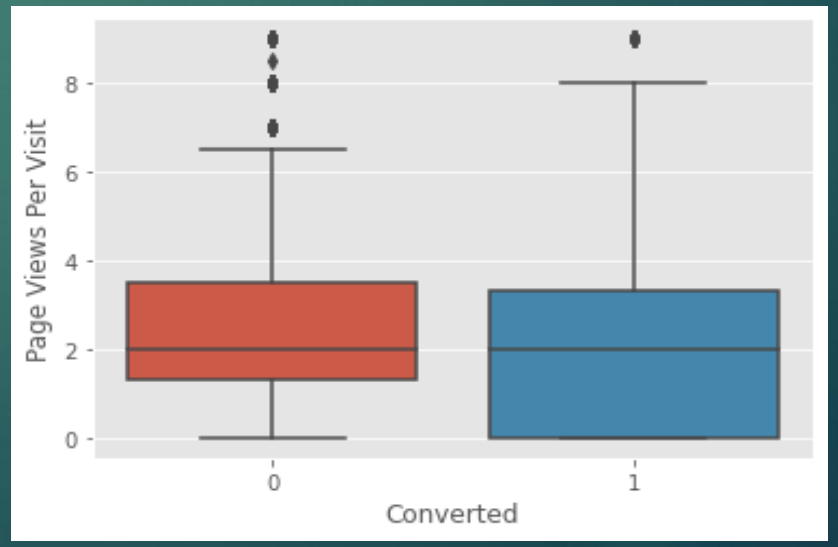
Total_views vs converted



Total_Time_spent on website vs Converted



Total Visits vs Converted



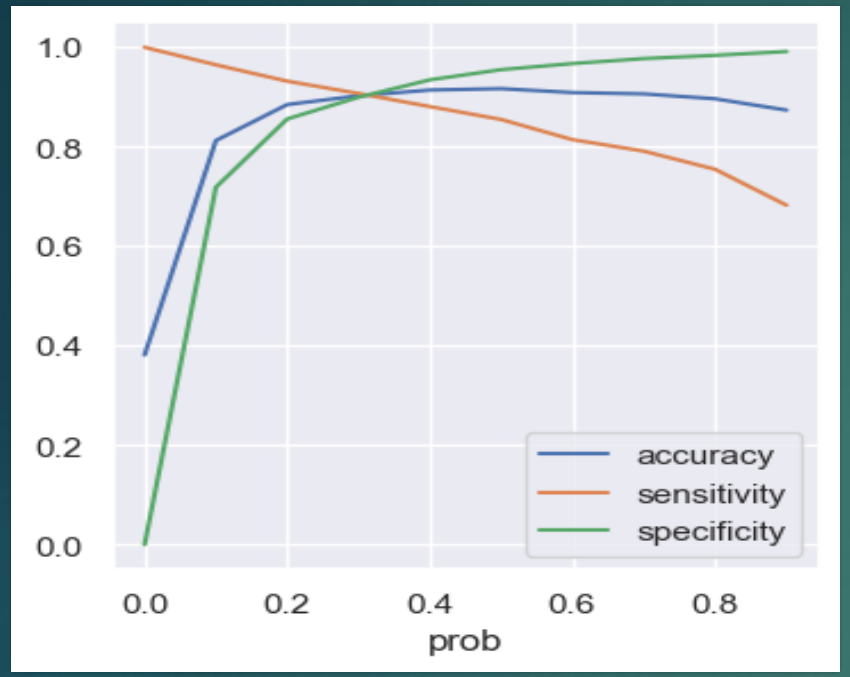
DATA CONVERSION:

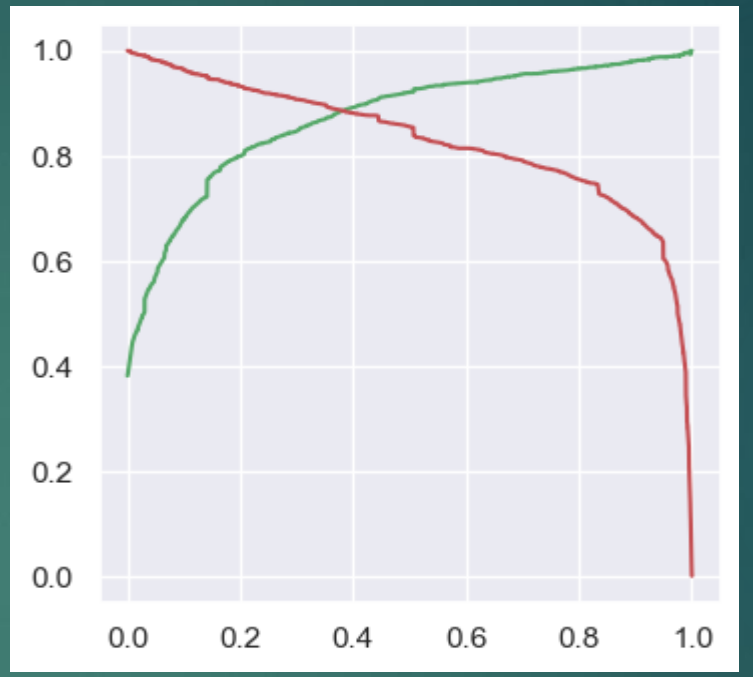
- Numerical variables are normalised.
- Dummy variables are created for objective type variables.
- Total Rows for Analysis = 9240
- Total columns for Analysis = 37

MODEL BUILDING

- Splitting the Data into Training and Testing Sets.
- The first basic step for regression is performing a train-test split , we have chosen 70:30 ratio.
- Use RFE for feature selection.
- Running RFE with 20 variables as output.
- Building Model by removing the variables whose p-value is greater than 0.05 and vif value is greater than 5.
- Predictions on test data set.
- Overall accuracy is 91%.

ROC CURVE





- Finding Optimal Cut off point.
- Optimal cut off probability is that where we get balanced sensitivity and specificity.
- From the first graph it is visible that the optimal cut off is at 0.35.

Conclusion:

- It was found that the variables that mattered the most in the potential buyers are (In descending order) :

1. What_matters_most_to_you_in_choosing_a_course
2. Tags_Will revert after reading the email
3. Last_Notable_Activity_Modified
4. a.Tags_Other Tags
   b.Tags_Ringing
5. a. Last_Activity_SMS Sent
   b. Last_Activity_Olark Chat Conversation
6. Occupation_Working Professional
7. Lead_Origin_Lead Import

- Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.