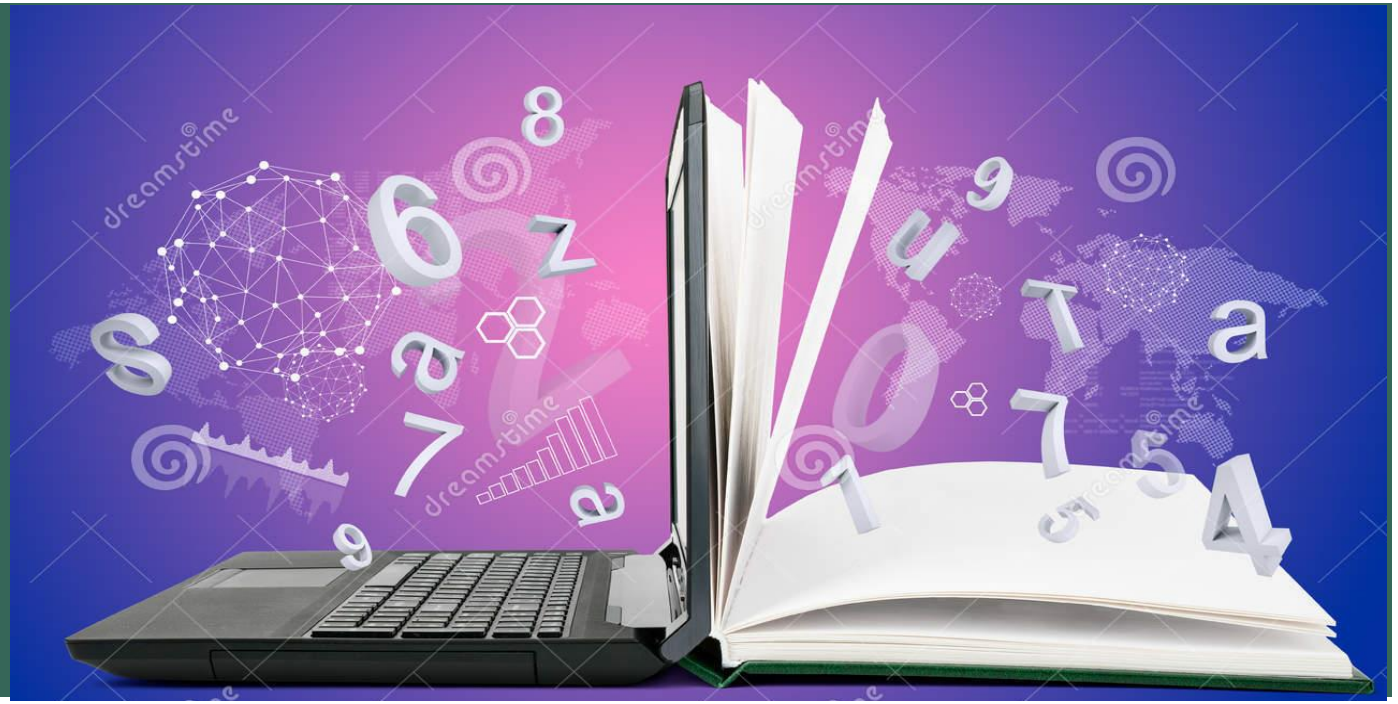
LEAD SCORING CASE STUDY

ANALYSED BY:

ANCHITA RANAKOTI ,

VISHWAKUMAR &

ASMITA COOMAR.



CONTENT

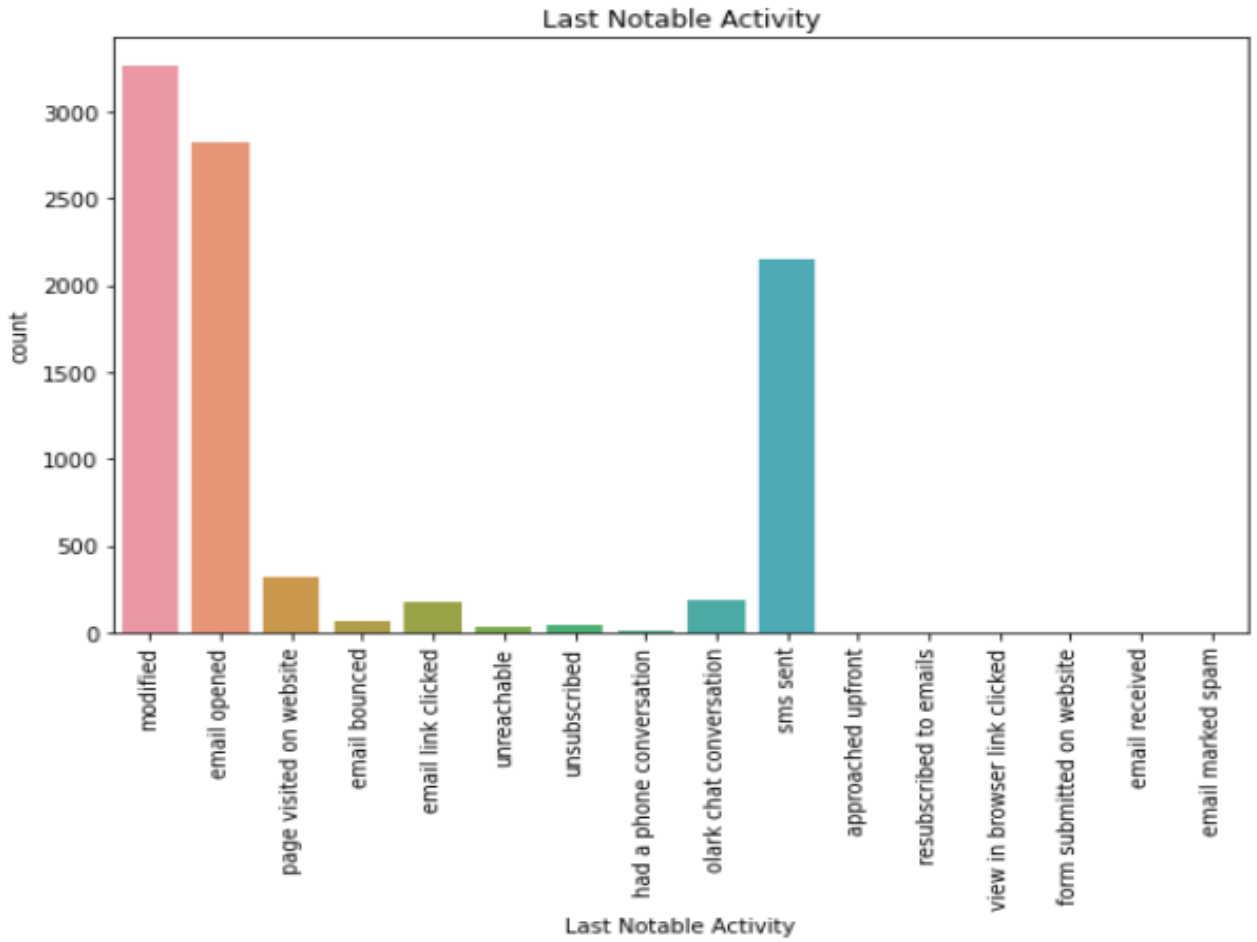
1. PROBLEM STATEMENT.
2. EXPLORATORY DATA ANALYSIS : UNIVARIATE, BIVARIATE & MULTIVARIATE ANALYSIS
3. MODEL BUILDING : FINAL VIF TABLE
4. MODEL EVALUATION : ROC CURVE AND ACCURACY, SENSITIVITY & SPECIFICITY CURVE
5. PRECISION & RECALL TRADE OFF
6. CONCLUSION

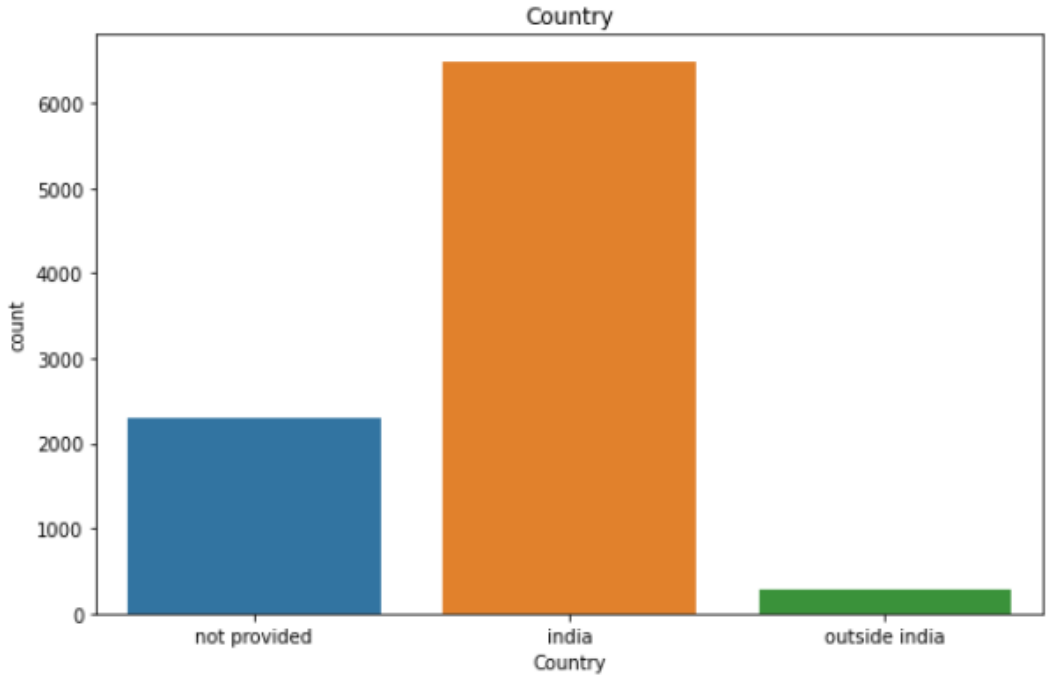
1. PROBLEM STATEMENT

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%.

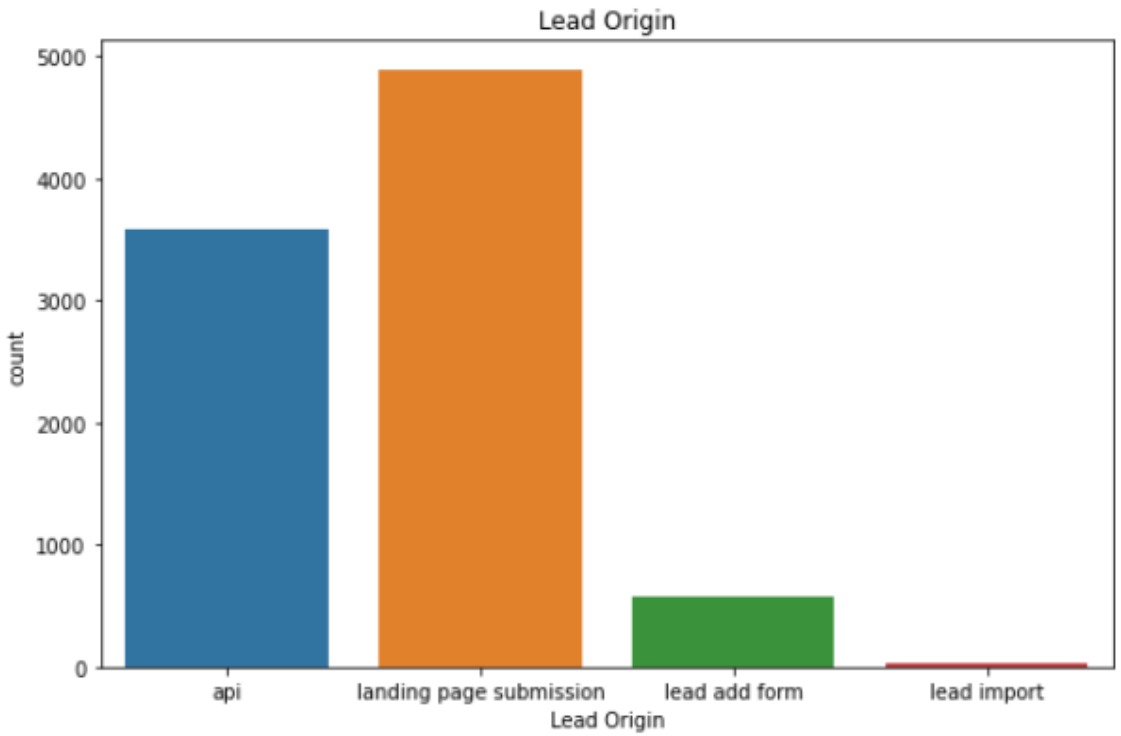
- **Goals of the Case Study:**
- To **build a logistic regression model to assign a lead score** between 0 and 100 to each of the leads which can be used by the company to target potential leads.
- To **adjust if the company's requirement changes** in the future so you will need to handle these as well.

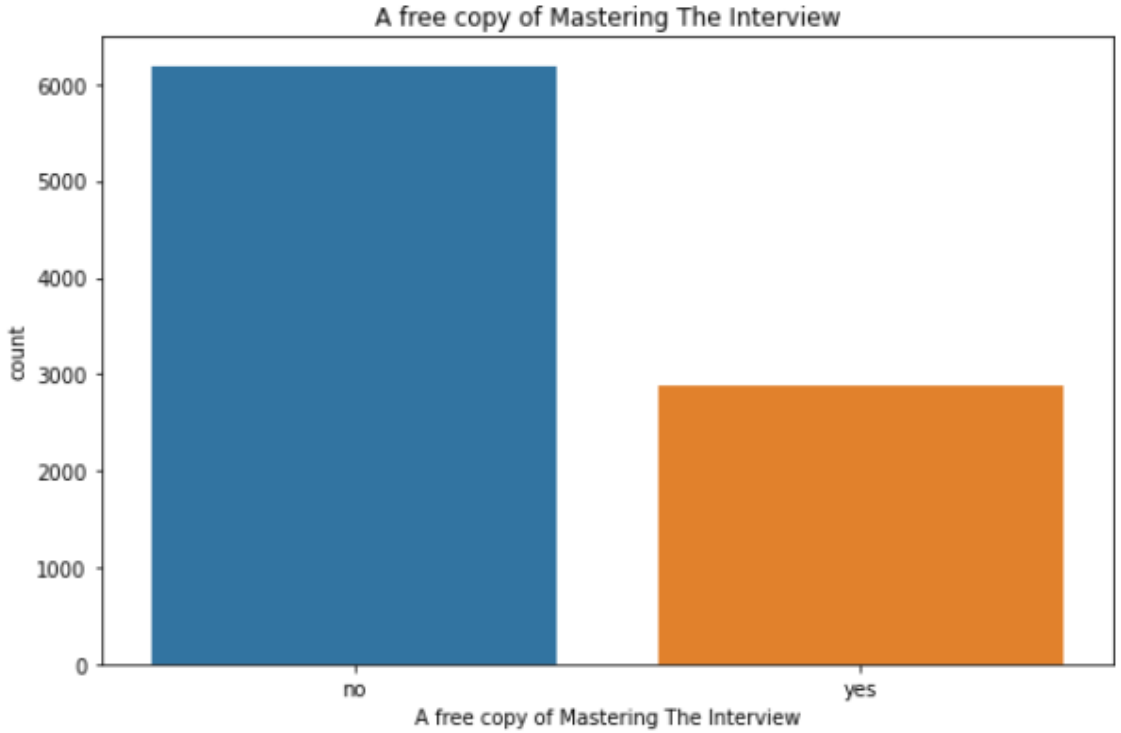
3. EXPLORATORY DATA ANALYSIS: UNIVARIATE ANALYSIS



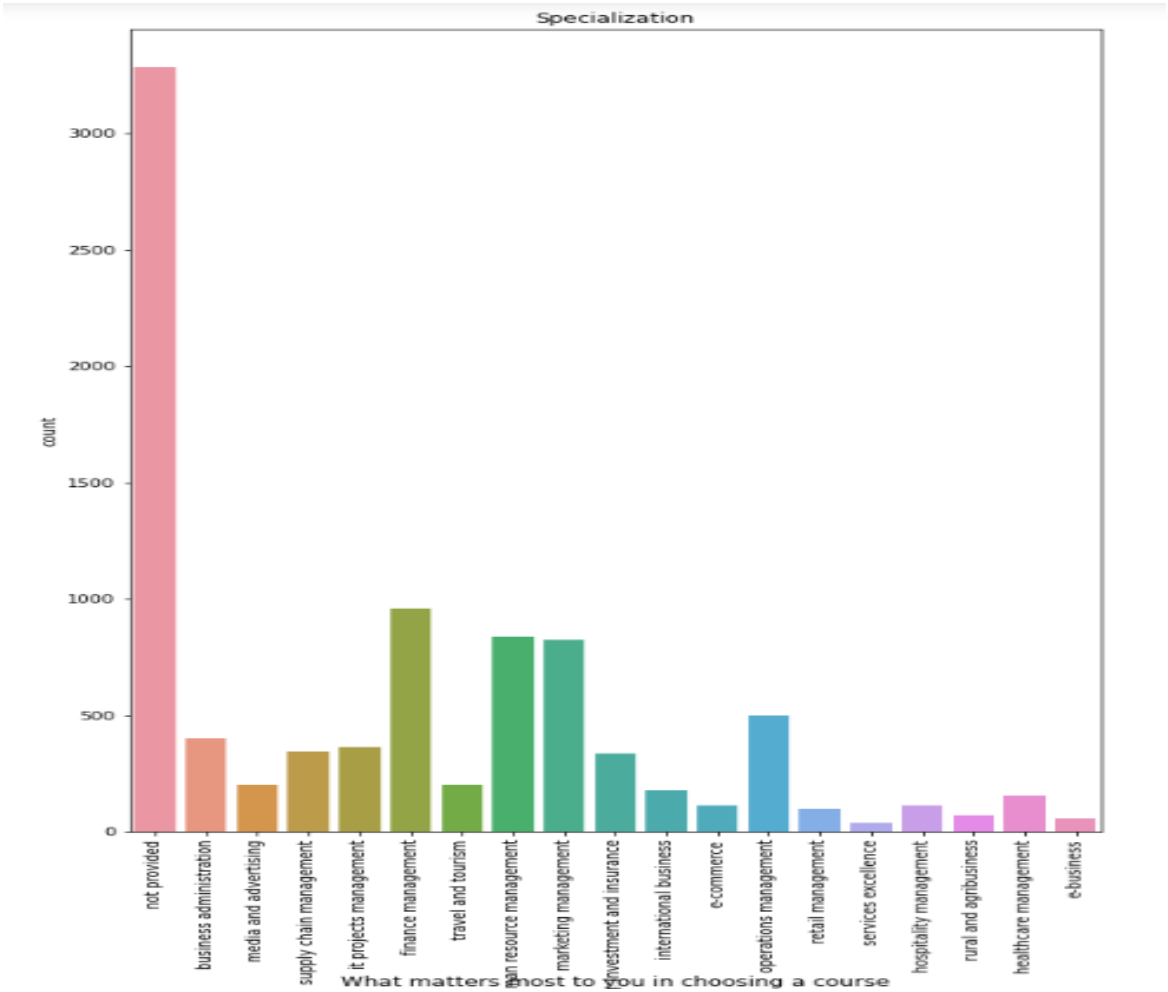


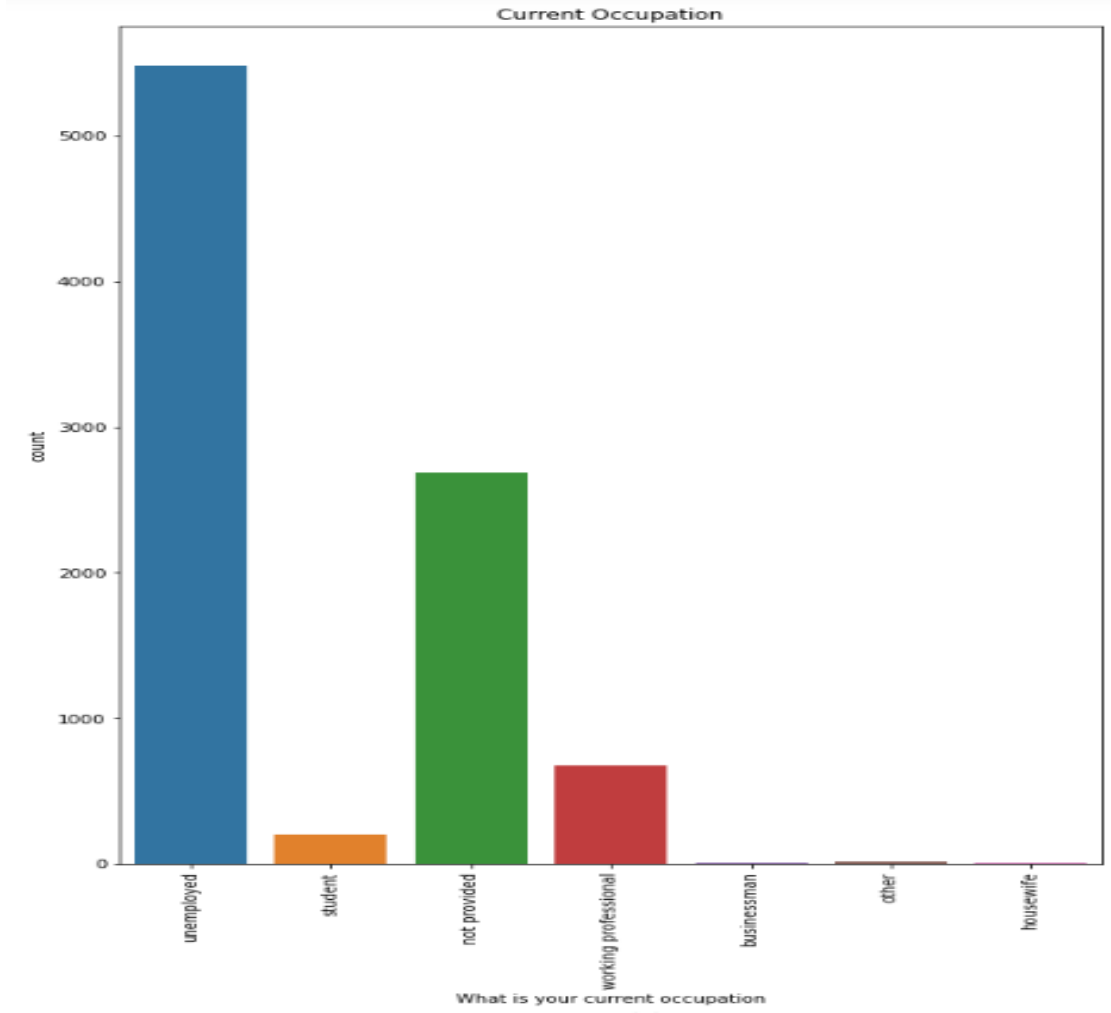
3. EXPLORATORY DATA ANALYSIS: UNIVARIATE ANALYSIS



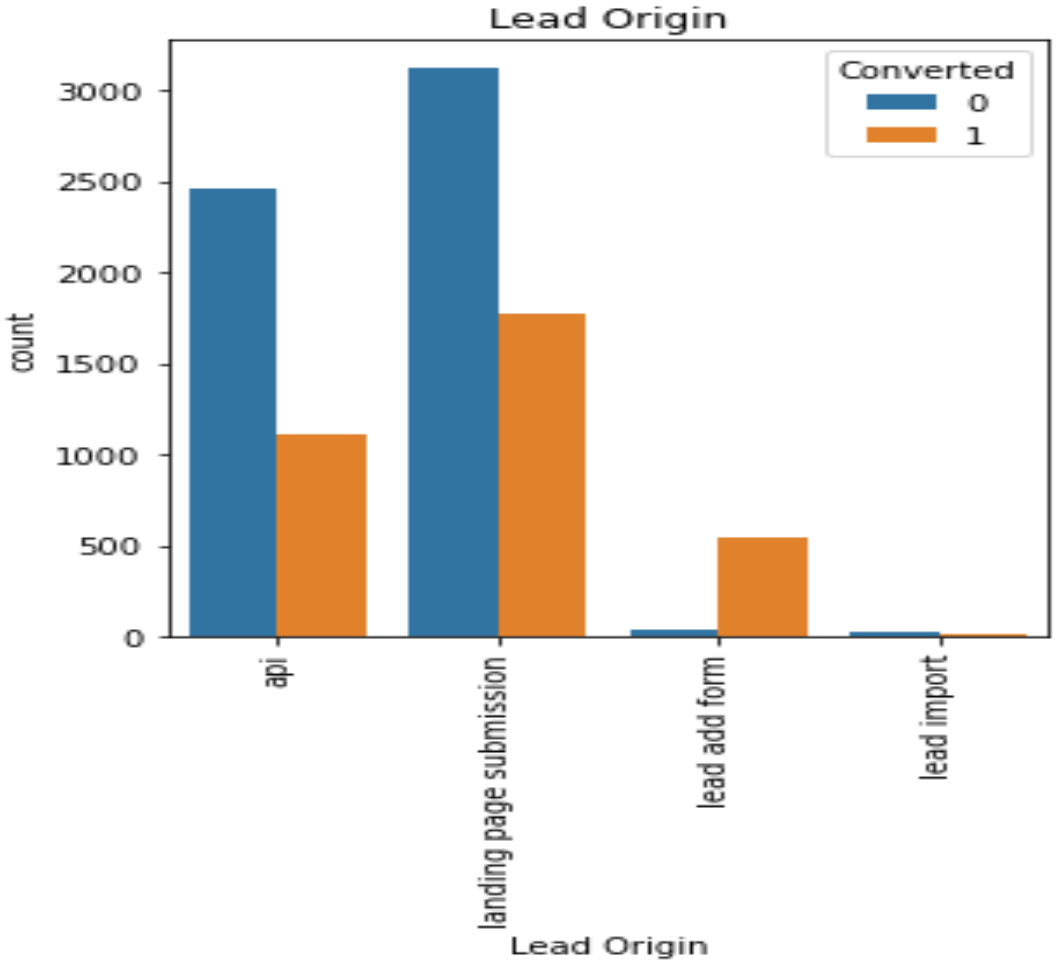


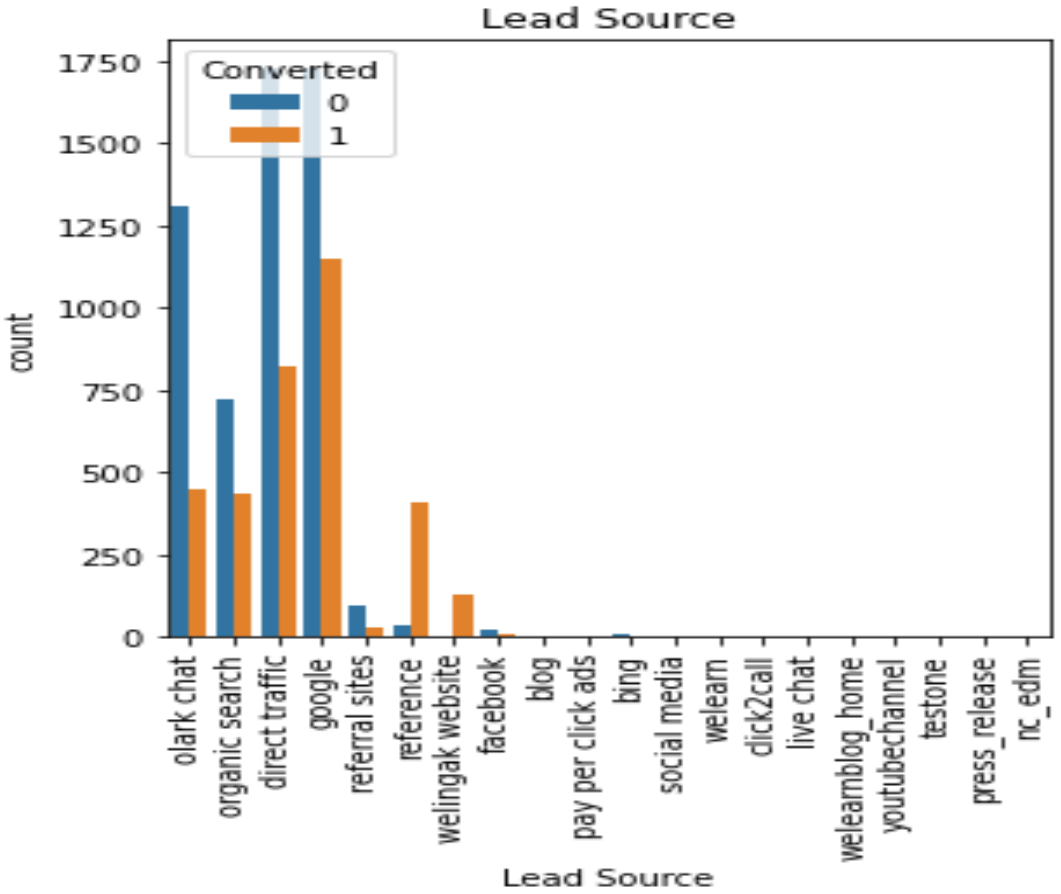
3. EXPLORATORY DATA ANALYSIS: BIVARIATE ANALYSIS



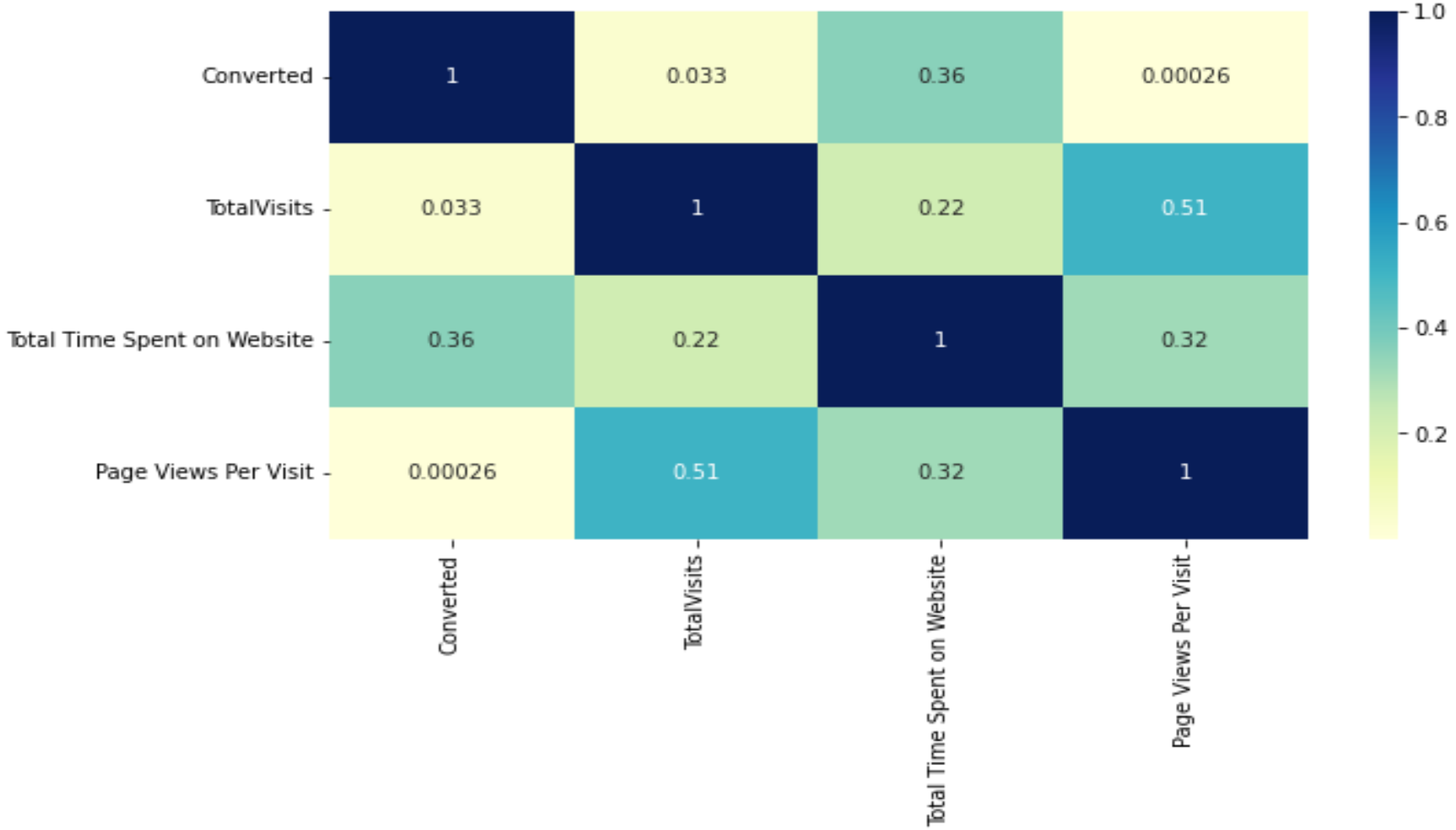


3. EXPLORATORY DATA ANALYSIS: BIVARIATE ANALYSIS





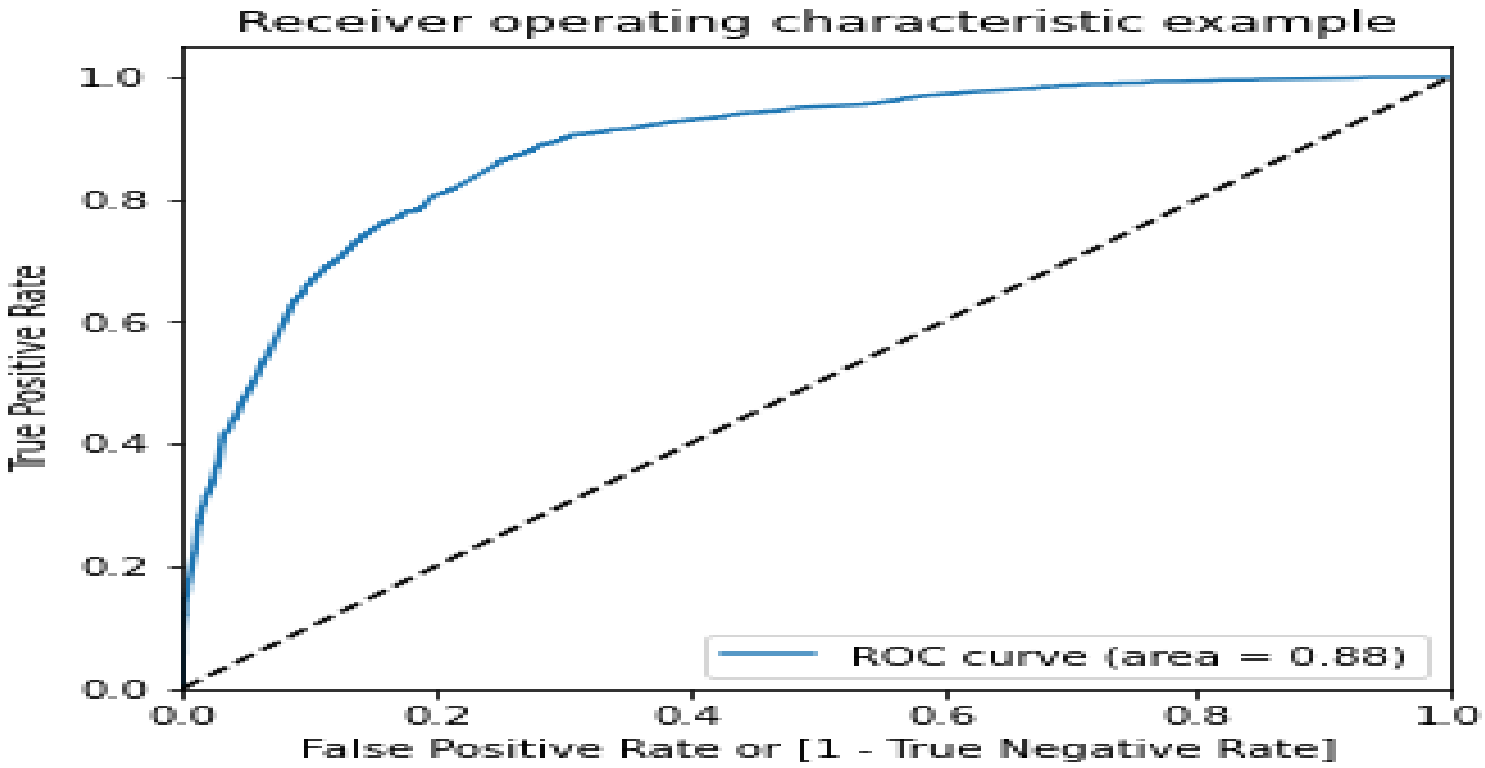
3. EXPLORATORY DATA ANALYSIS: MULTIVARIATE ANALYSIS



4. MODEL BUILDING: FINAL VIF TABLE

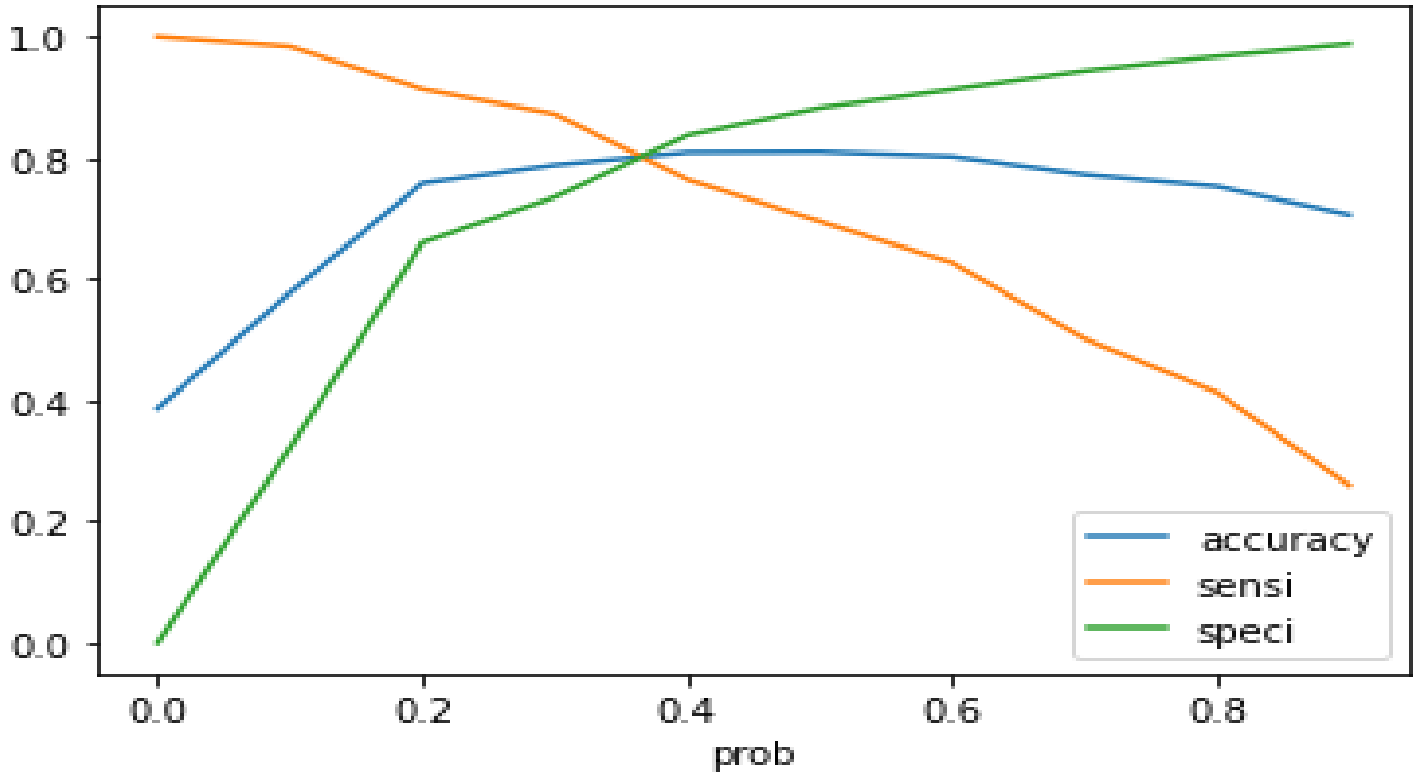
Features	VIF	
9	What is your current occupation_unemployed	2.29
1	Total Time Spent on Website	2.06
0	TotalVisits	1.84
2	Lead Origin_lead add form	1.58
7	Last Activity_sms sent	1.53
3	Lead Source_olark chat	1.51
6	Last Activity_olark chat conversation	1.37
10	What is your current occupation_working profes...	1.32
4	Lead Source_welingak website	1.31
5	Do Not Email_yes	1.06
8	What is your current occupation_student	1.05
11	Last Notable Activity_unreachable	1.01

5. MODEL EVALUATION: ROC CURVE



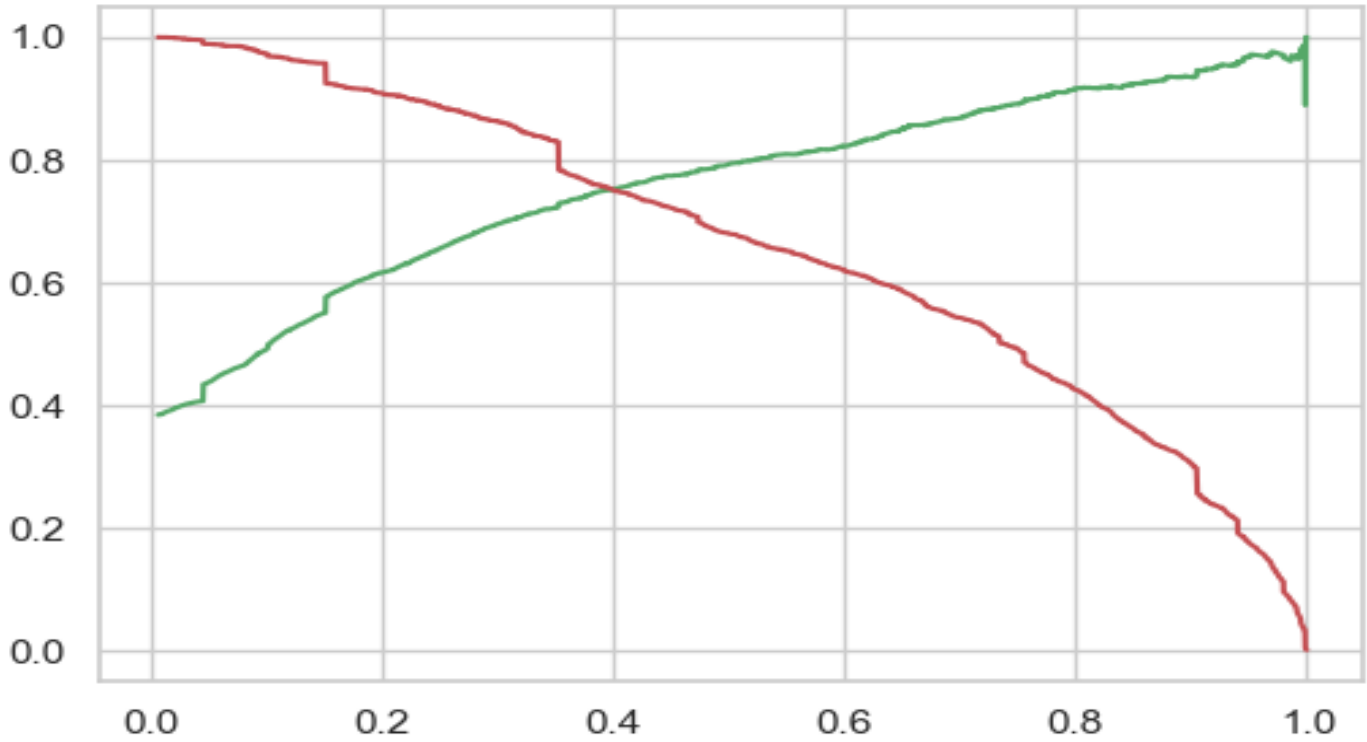
The area under ROC curve is 0.88 which is a very good value.

5. MODEL EVALUATION: ACCURACY, SENSITIVITY & SPECIFICITY CURVE



From the graph it is visible that the optimal cut off is at 0.35.

6. PRECISION & RECALL TRADE OFF



With the current cut off as 0.41 we have Precision around 74% and Recall around 76%

7. CONCLUSION

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

1. The total time spend on the Website.
2. Total number of visits.
3. When the lead source was:
   a. Google
   b. Direct traffic
   c. Organic search
   d. Welingak website
4. When the last activity was:
   a. SMS
   b. Olark chat conversation
5. When the lead origin is Lead add format.
6. When their current occupation is as a working professional.

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.


Thank You