

Lead Score Case Study

Prepared By

T Sai Vishwanath

Problem Statement

- ▶ Industry professionals can purchase online courses from X Education..
- ▶ Despite receiving a large number of leads, X Education has a relatively low lead conversion rate. For instance, only approximately 30 of 100 leads they could gather in a day might actually be converted.
- ▶ To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
- ▶ The lead conversion rate should increase if they are successful in locating this group of leads because the sales staff will be spending more time speaking with potential leads rather than calling everyone.

Business Objective:

- ▶ X education wants to know most promising leads.
- ▶ They want to develop a model that detects the hot leads for that..
- ▶ Deployment of the model for the future use.

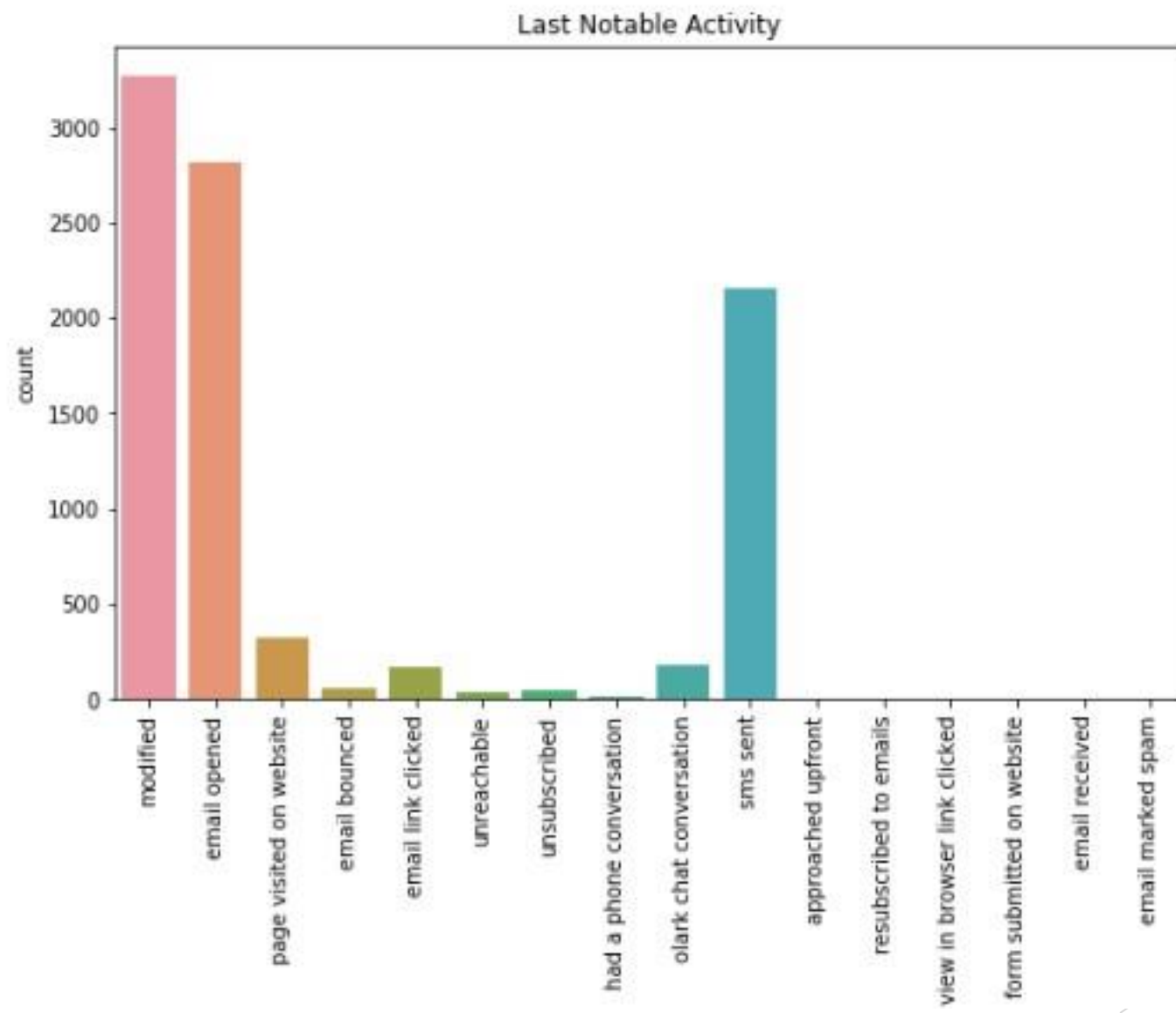
Solution Methodology

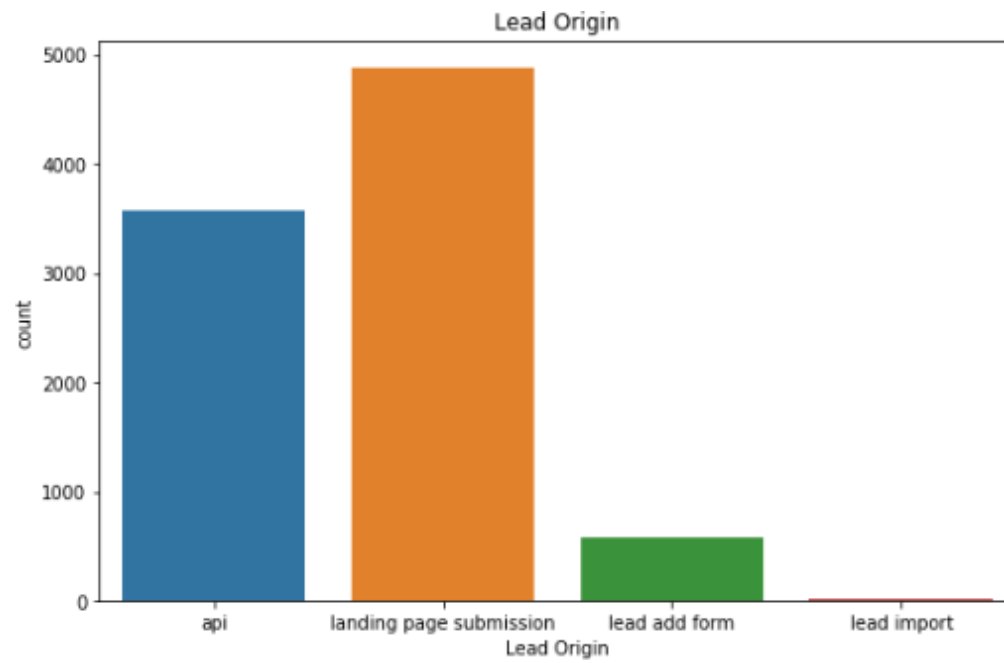
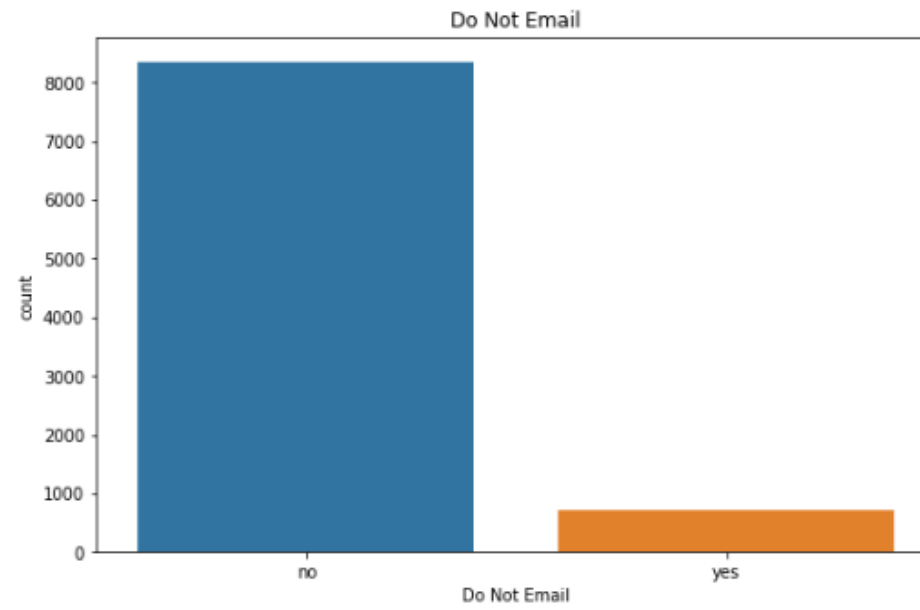
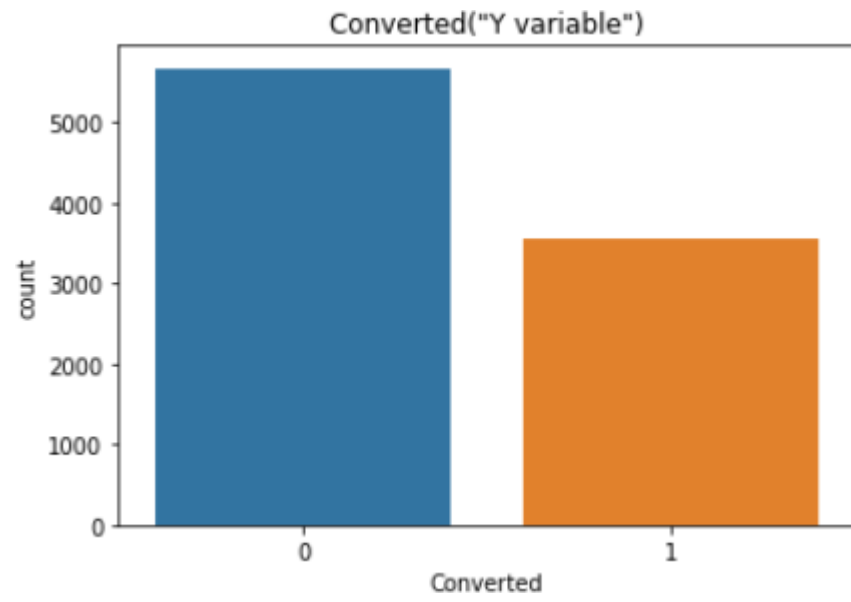
- ▶ Data cleaning and data manipulation.
 1. Verify and deal with duplicate data.
 2. Verify and manage missing values and NA values.
 3. Exclude columns from the analysis if they have a significant number of missing values..
 4. Imputation of the values, if necessary.
 5. Check and handle outliers in data.
- ▶ EDA
 1. Univariate data analysis: value counts, variable distributions, etc..
 2. Bivariate data analysis: patterns between the variables and correlation coefficients, etc.
- ▶ Data encoding, feature scaling, and dummy variables.
- ▶ Using logistic regression as a classification tool, make predictions and build models..
- ▶ Validation of the model.
- ▶ Model presentation.
- ▶ Conclusions and recommendations.

Data Manipulation

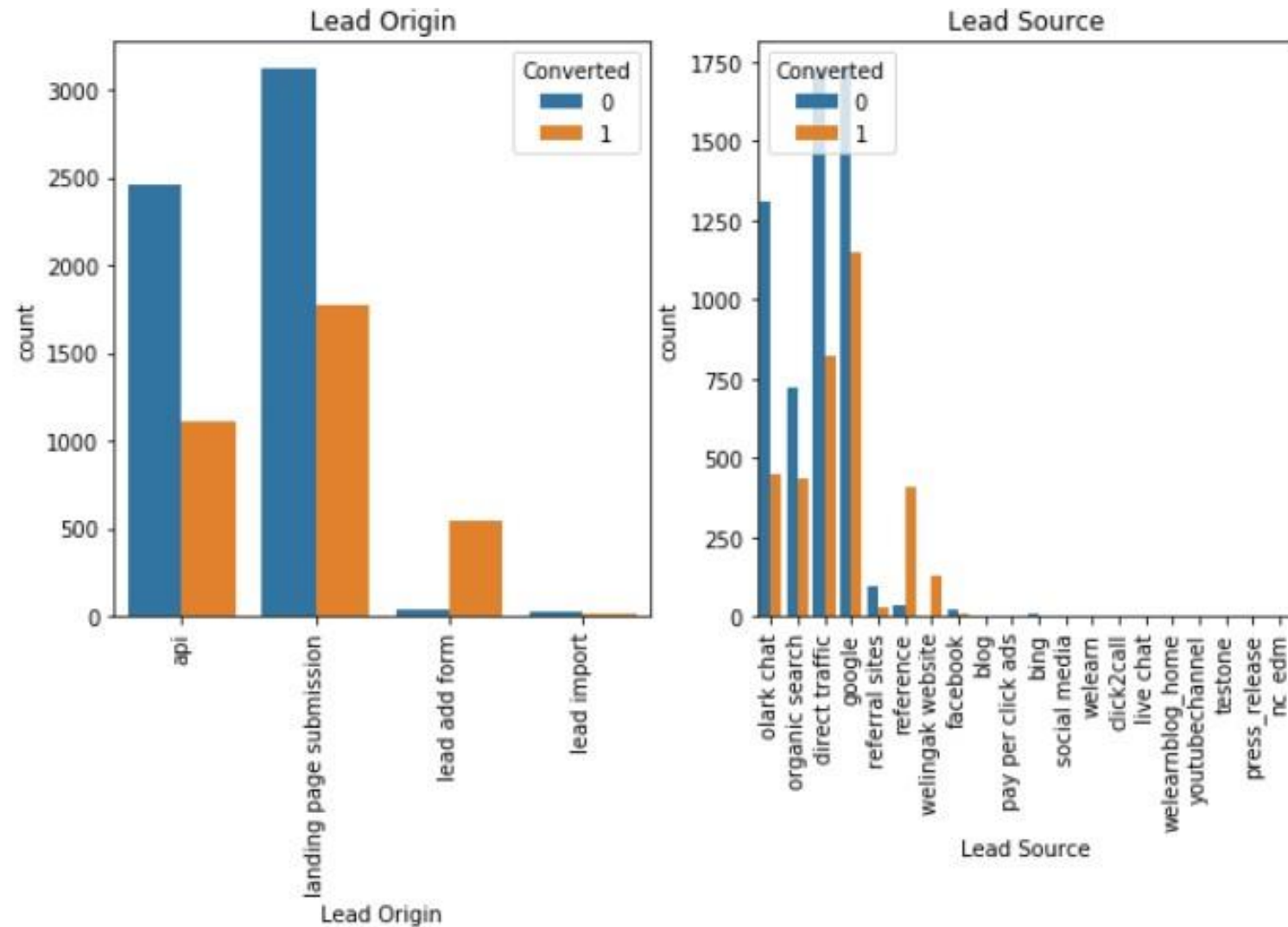
- ▶ Total Number of Rows =37, Total Number of Columns =9240.
- ▶ Features with a single purpose, such as "Magazine," "Get More Information About Our Courses," and "Update me on Supply"
- ▶ Chain Content", "Get updates on DM Content", "I agree to pay the amount through cheque" etc. have been dropped.
- ▶ Removing the "Lead Number" and "Prospect ID" that are not required for the analysis.
- ▶ We have discontinued various features after examining the value counts for some of the object type variables. These features include "Do Not Call," "What factors are most important to you while choosing a course," "Search," "Newspaper Article," "X Education Forums," "Newspaper," "Digital Advertising," etc.
- ▶ Removing the columns like "How did you hear about X Education" and "Lead Profile" that had more than 35% of their values missing.

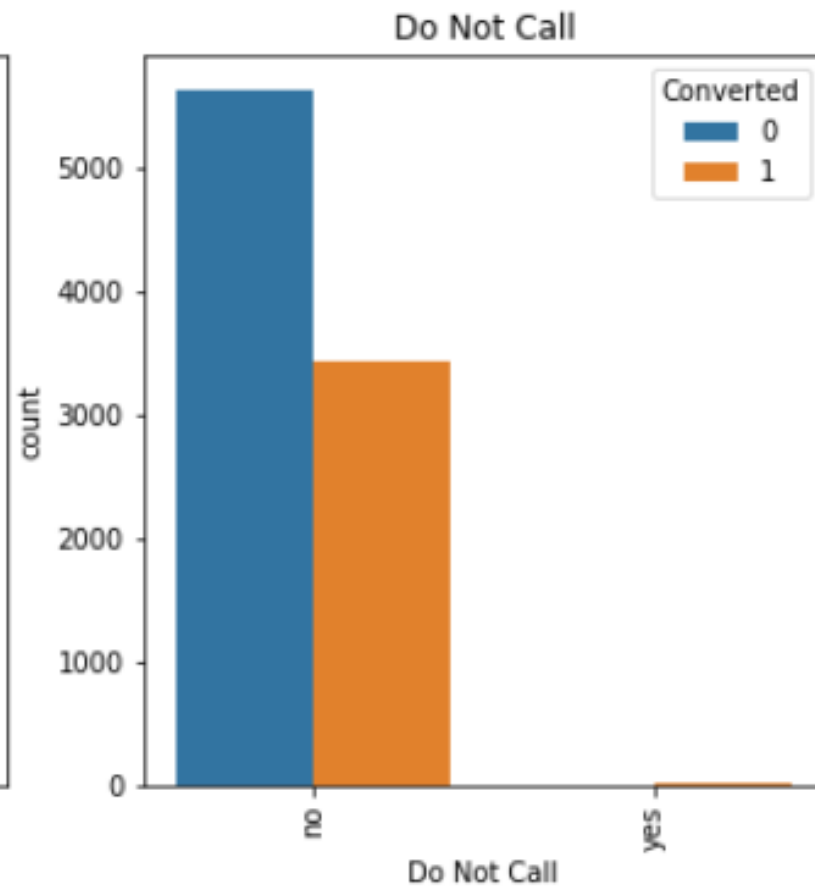
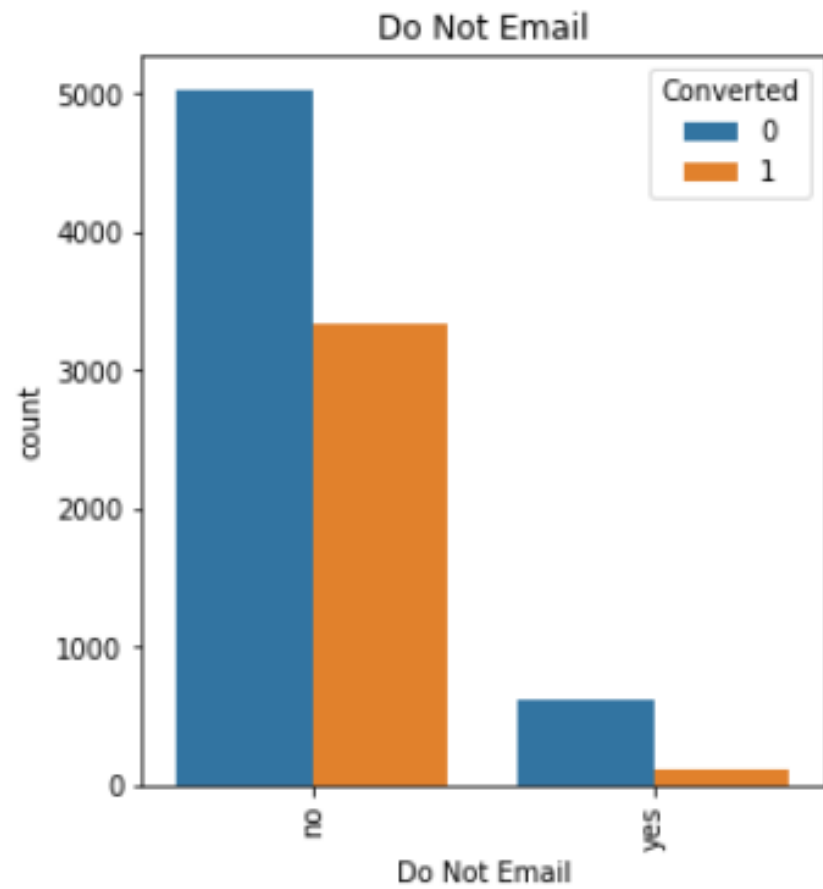
EDA

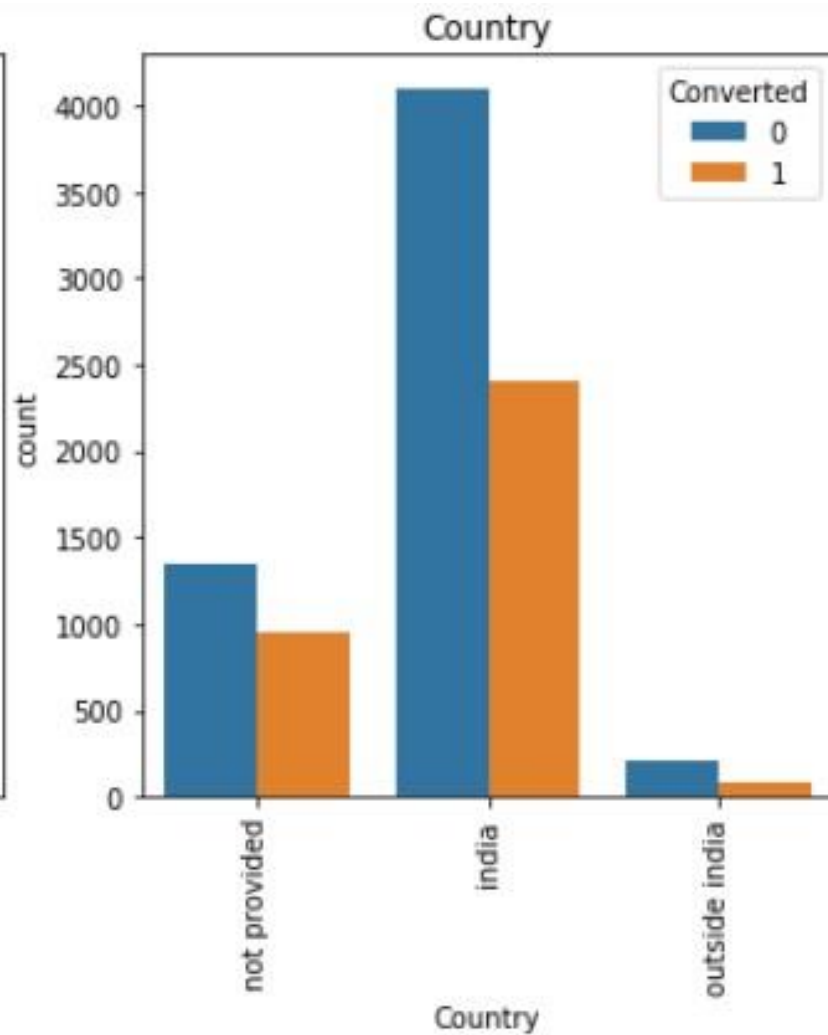
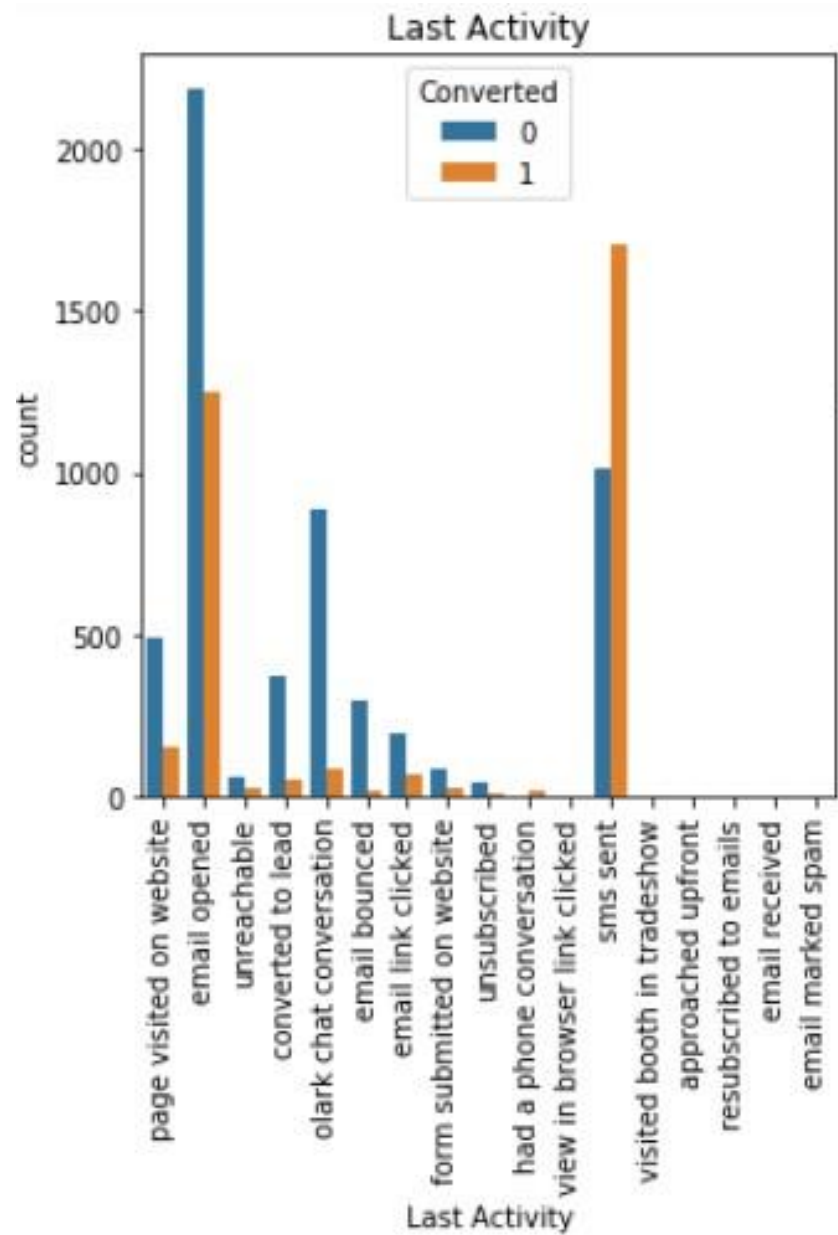




Categorical Variable Relation







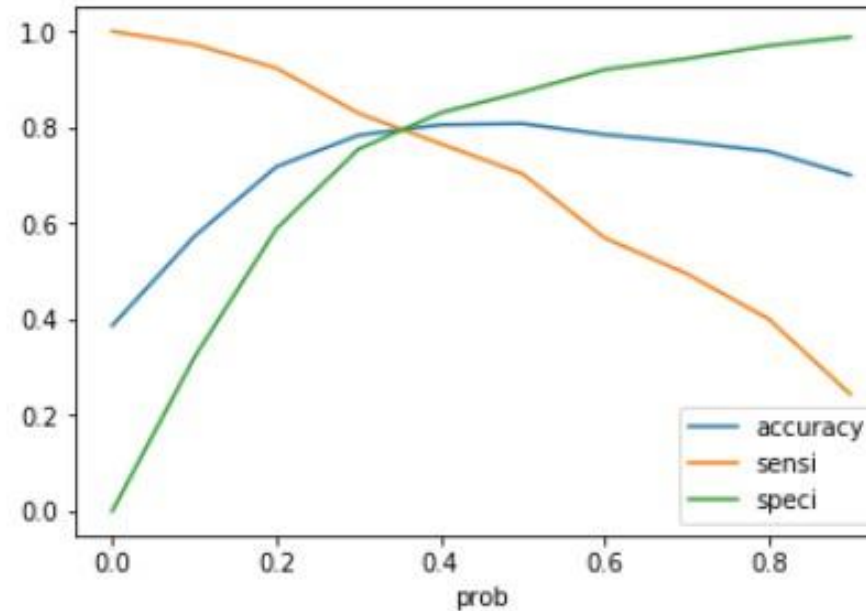
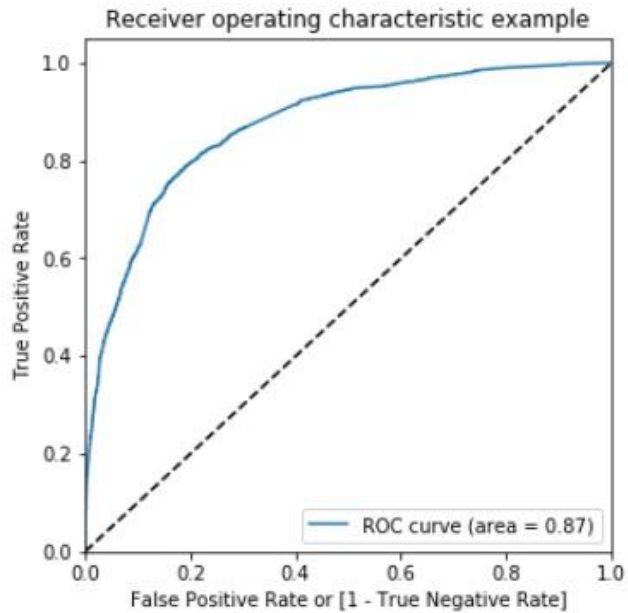
Data Conversion

- ▶ Numerical Variables are Normalised
- ▶ Dummy Variables are created for object type variables
- ▶ Number of total Rows for Analysis: 8792
- ▶ Number of total Columns for Analysis: 43

Model Building

- ▶ Splitting the Data into Training and Testing Sets
- ▶ A train-test split is the first fundamental stage in the regression process; we have selected a 70:30 split.
- ▶ Use RFE for Feature Selection
- ▶ Running RFE with 15 variables as output
- ▶ Removing variables from the model whose p-value is higher than 0.05 and vif value is higher than 5
- ▶ Predictions on test data set
- ▶ Overall accuracy 81%

ROC Curve



- ▶ Finding Optimal Cut off Point
- ▶ The probability with the best sensitivity and specificity is known as the optimal cut off probability.
- ▶ The second graph makes clear that 0.35 is the ideal cut off..

Conclusion

According to research, the factors that affected potential customers the most were (In descending order) :

- ▶ The total time spend on the Website.
 - ▶ Total number of visits.
 - ▶ When the lead source was:
 - a. Google
 - b. Direct traffic
 - c. Organic search
 - d. Welingak website
 - ▶ When the last activity was:
 - a. SMS
 - b. Olark chat conversation
 - ▶ When the lead origin is Lead add format.
 - ▶ When their current occupation is as a working professional.
- With these in mind, X Education can succeed since they have a very good probability of persuading nearly all prospective customers to change their minds and purchase their courses.