

Lead Scoring Case Study

SUBMITTED BY-

SHILPA JOSHI, SRUTHI SHYAMALA
PRADEEP M

Lead Score Case Study for X Education

Problem Statement :

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

Business Objective

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

Methodology

- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.
- Conclusion

Problem solving methodology

Data Sourcing , Cleaning and Preparation

- Read the Data from Source
- Convert data into clean format suitable for analysis
- Remove duplicate data
- Outlier Treatment
- Exploratory Data Analysis
- Feature Standardization.

Feature Scaling and Splitting Train and Test Sets

- Feature Scaling of Numeric data
- Splitting data into train and test set.

Problem solving methodology

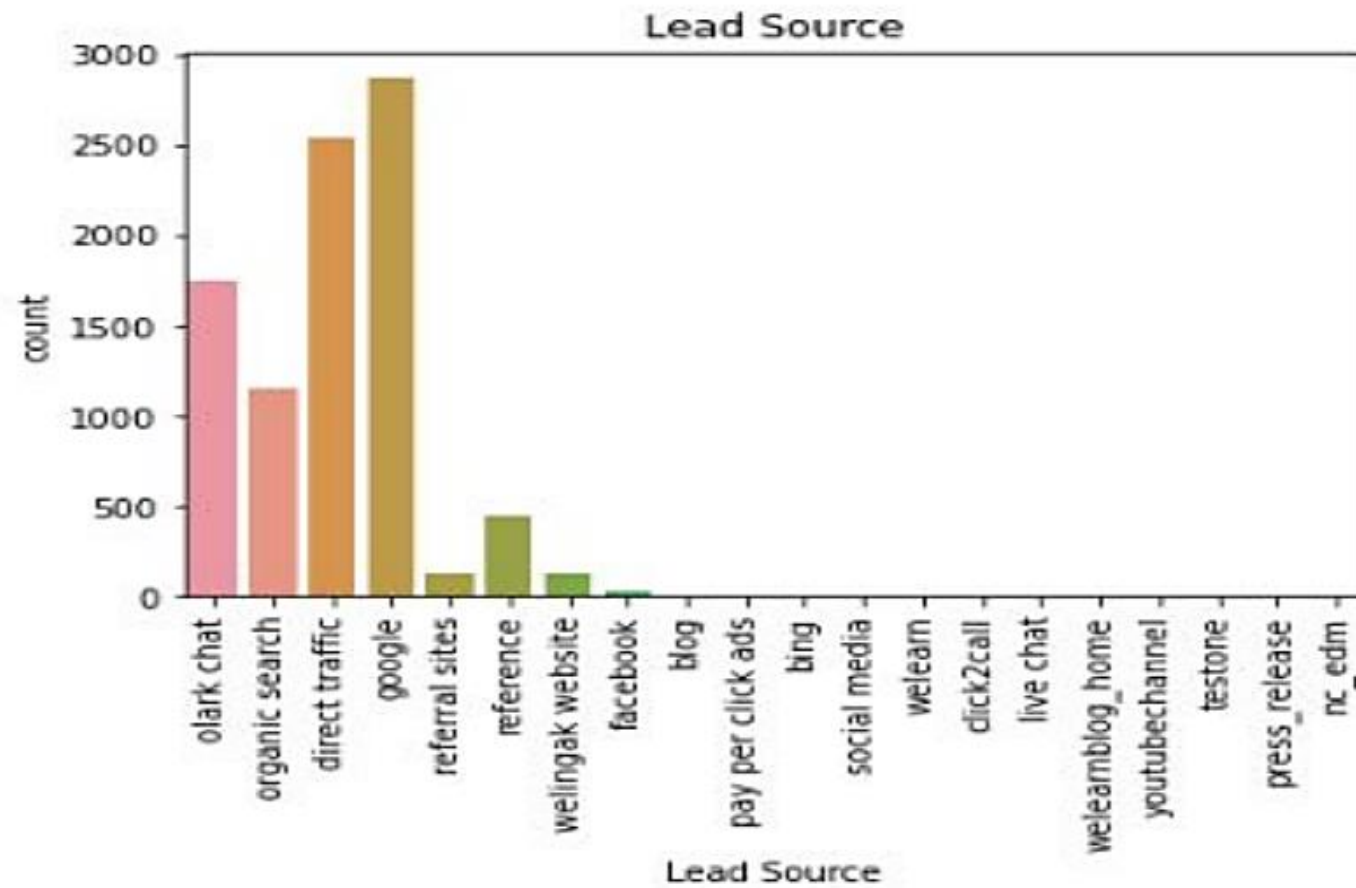
Model Building

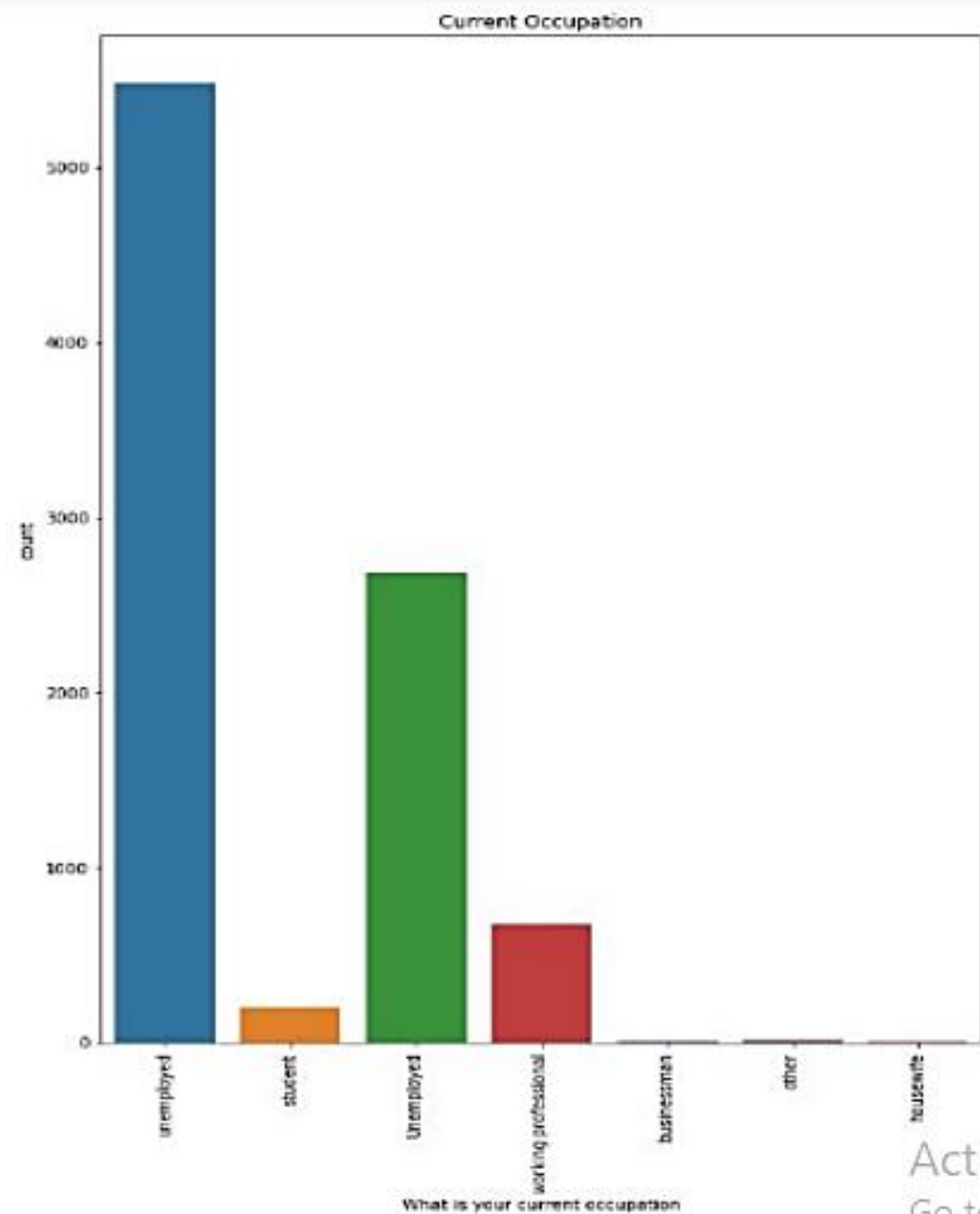
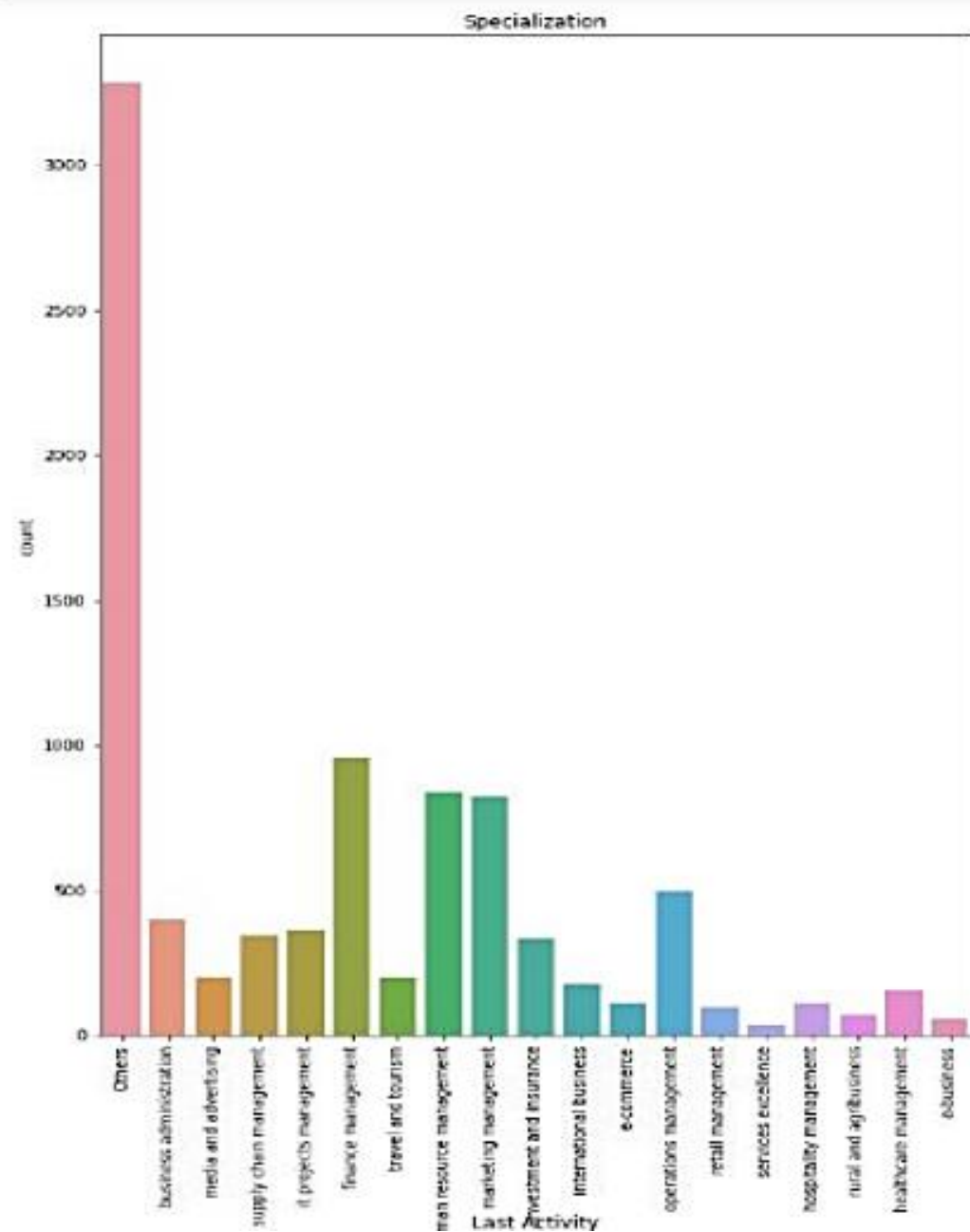
- Feature Selection using RFE
- Determine the optimal model using Logistic Regression
- Calculate various metrics like accuracy, sensitivity, specificity, precision and recall and evaluate the model.

Result

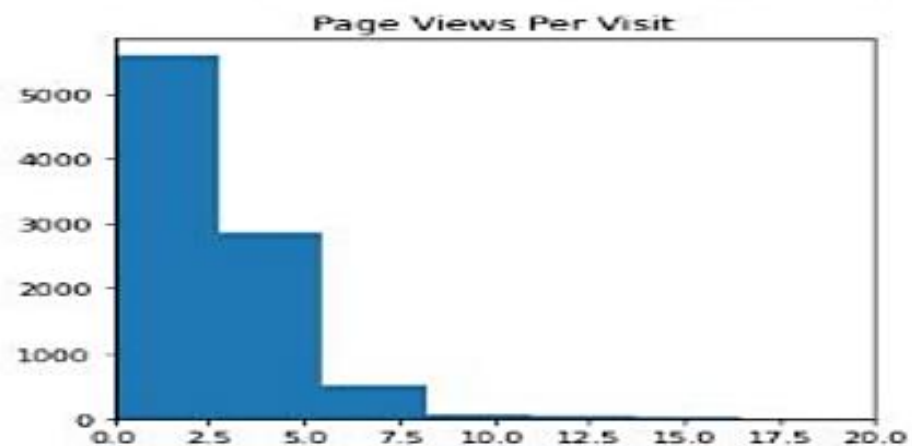
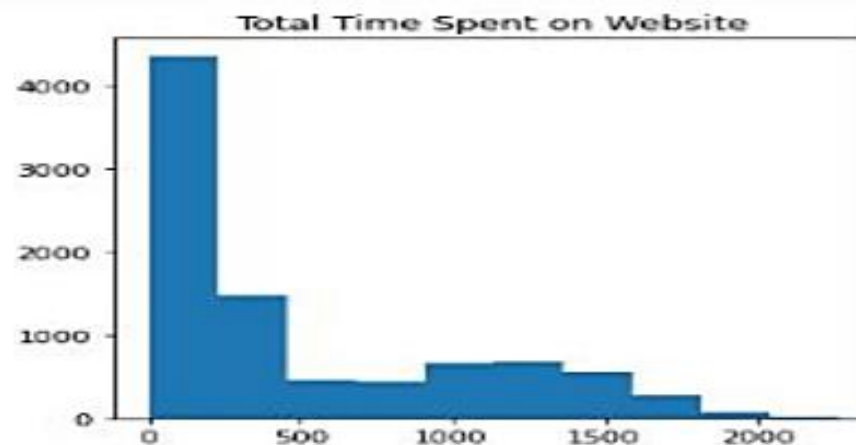
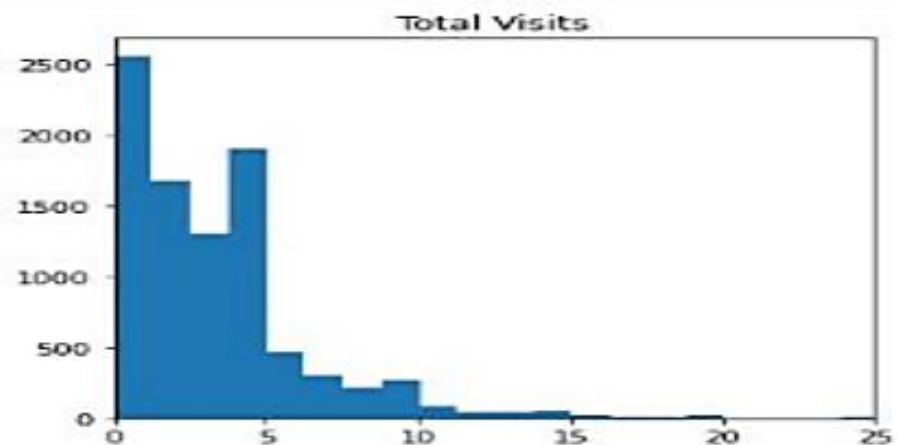
- Determine the lead score and check if target final predictions amounts to 80% conversion rate.
- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics

EDA

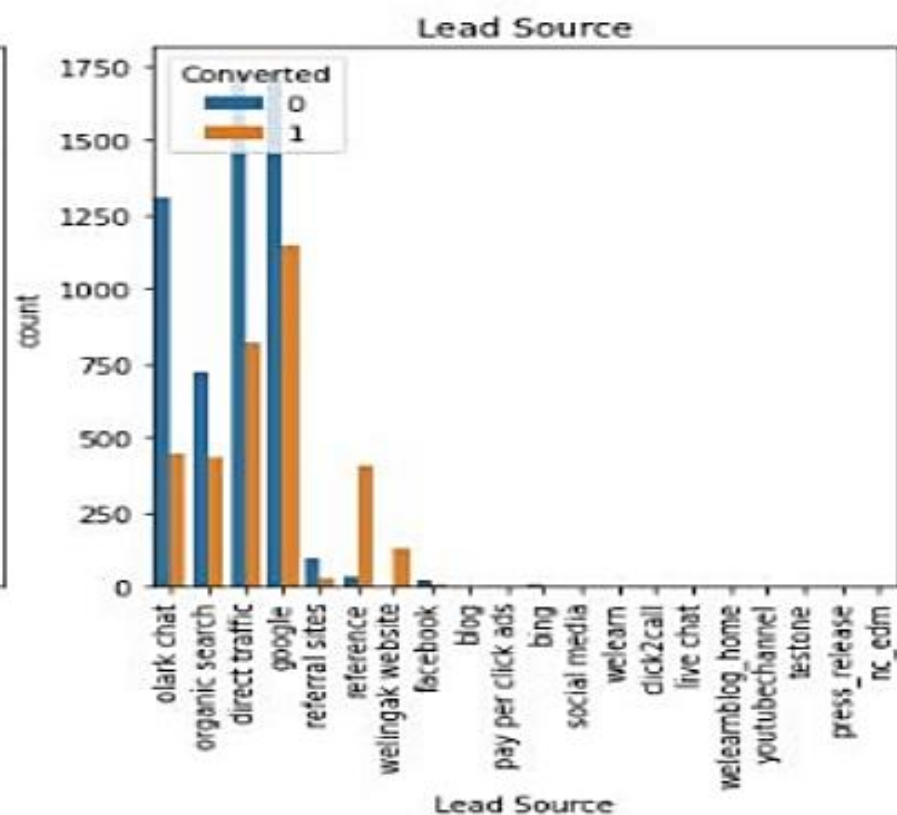
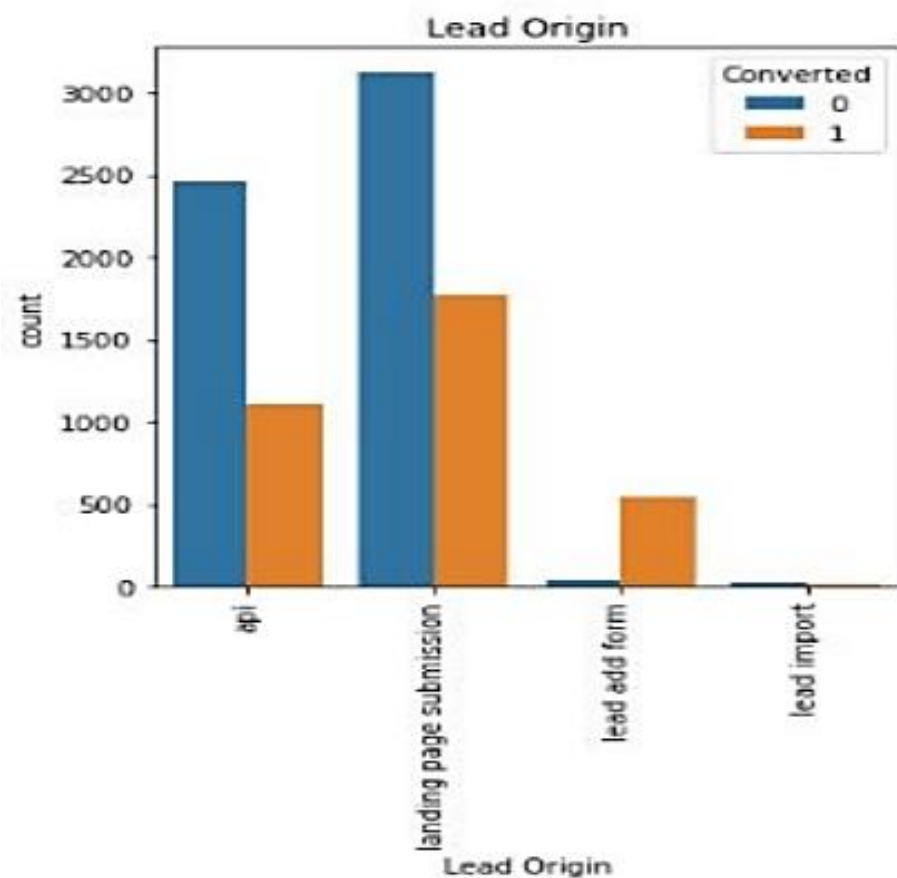




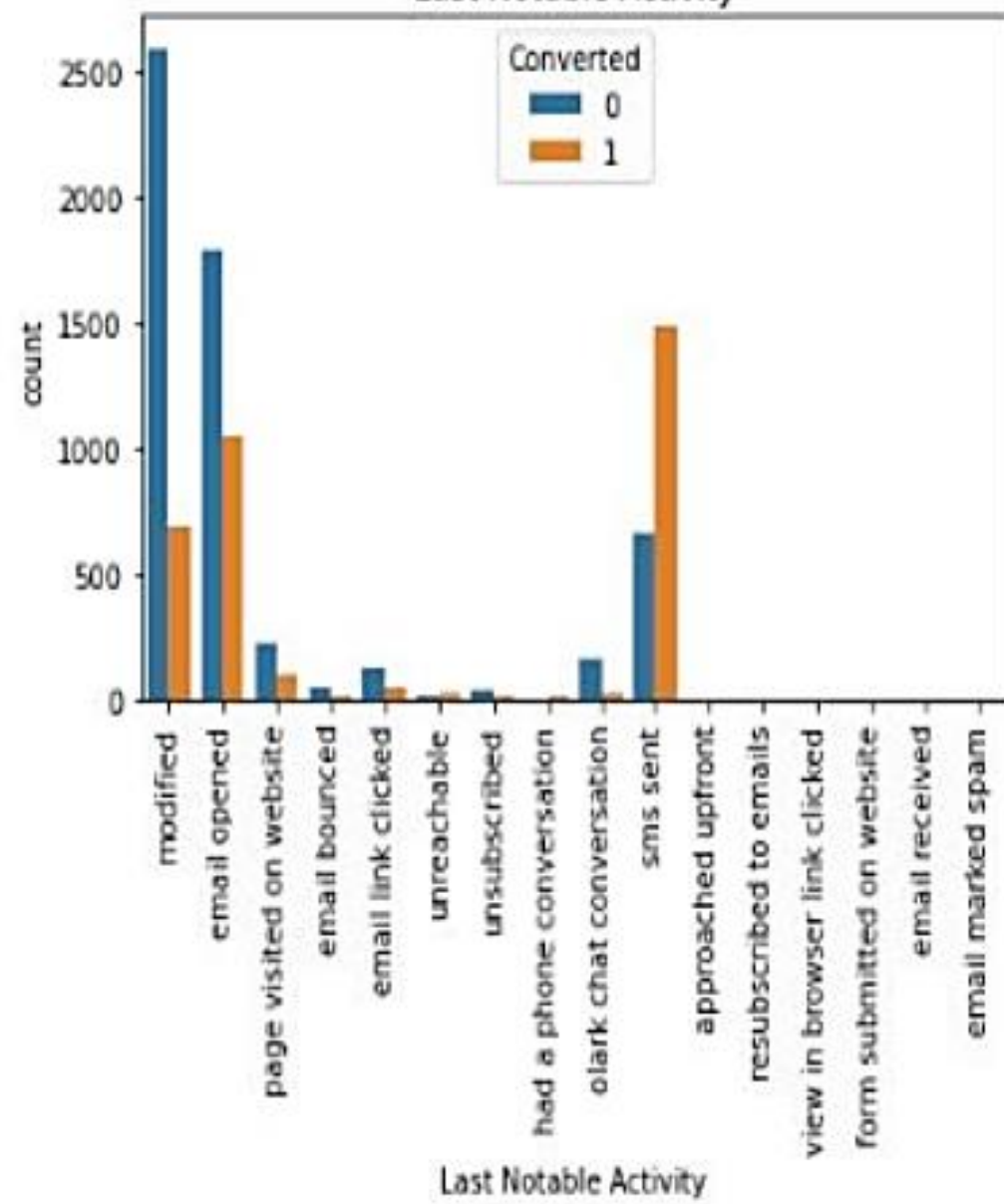
Numerical Variables



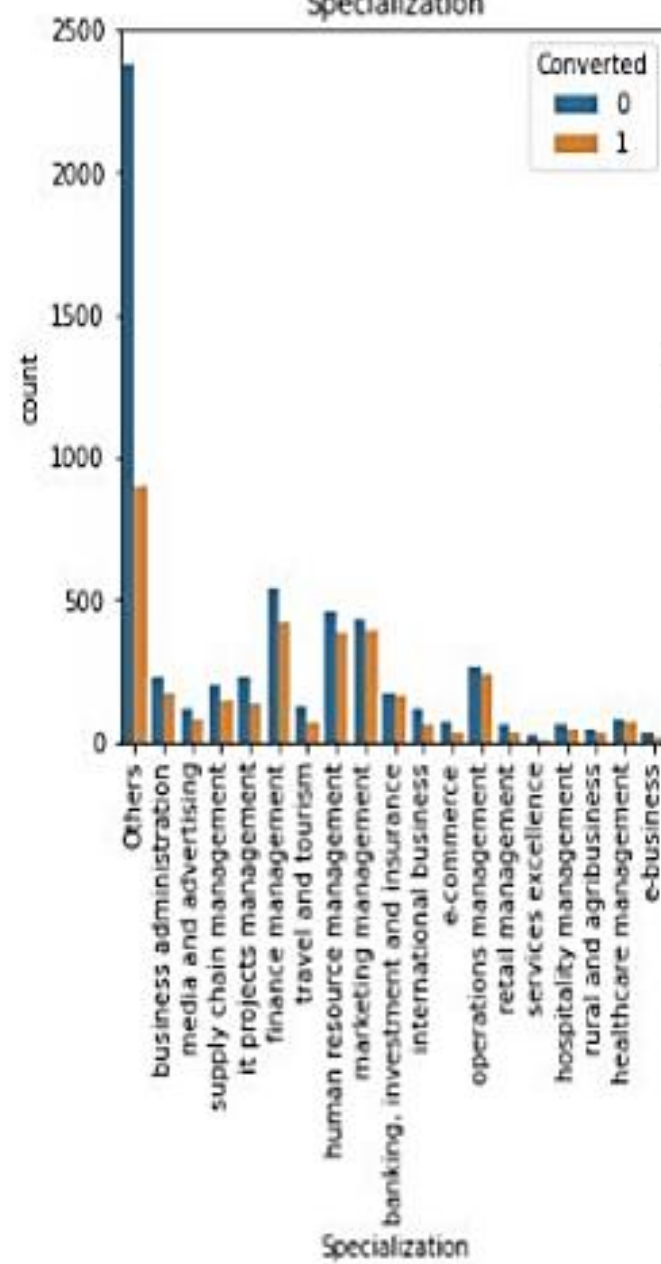
Categorical Variables



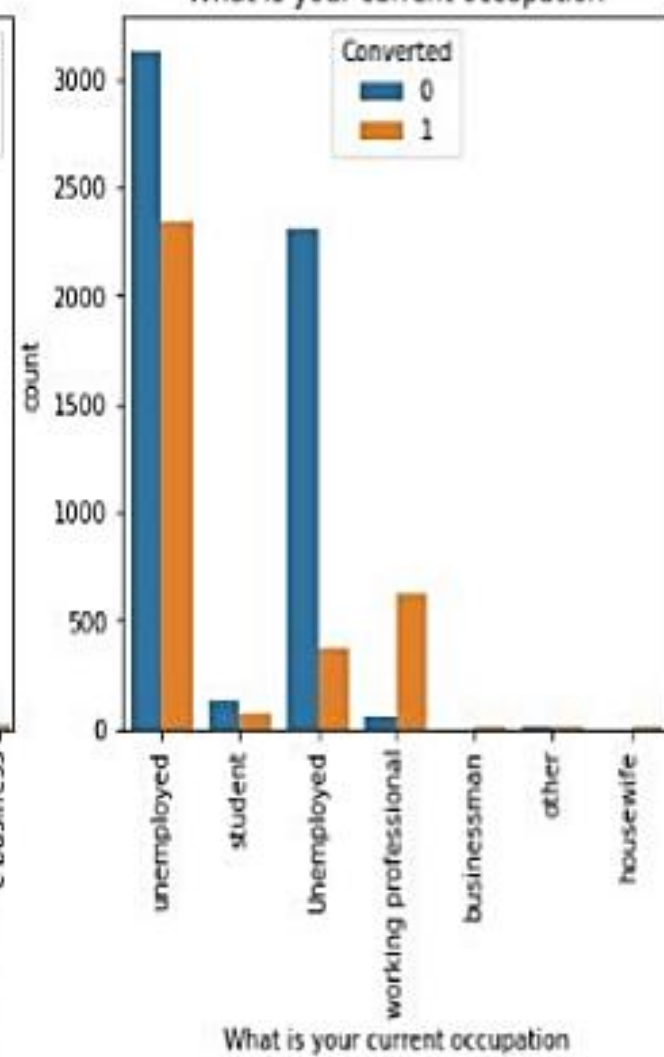
Last Notable Activity



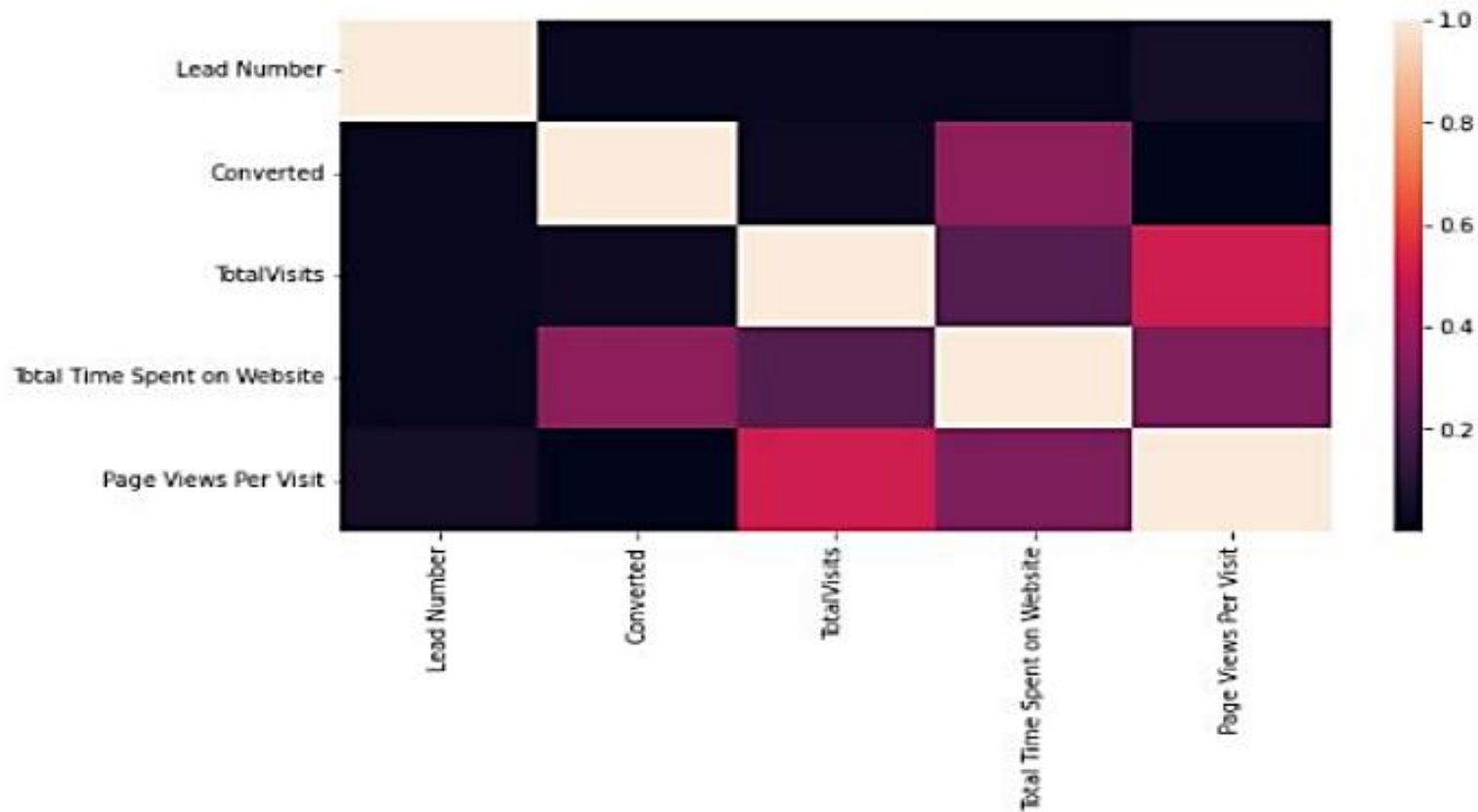
Specialization



What is your current occupation



Correlation



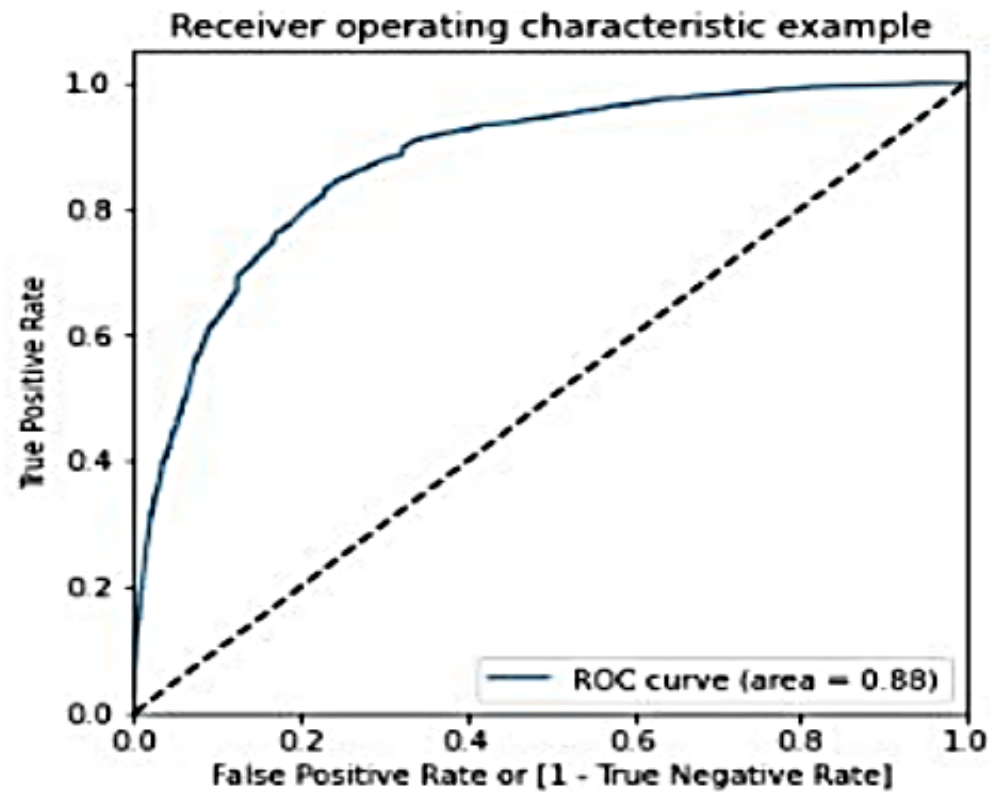
Data Conversion

- Numerical Variables are Normalised
- Dummy Variables are created for object type variables
- Total Rows for Analysis: 9240
- Total Columns for Analysis: 37

Model Building

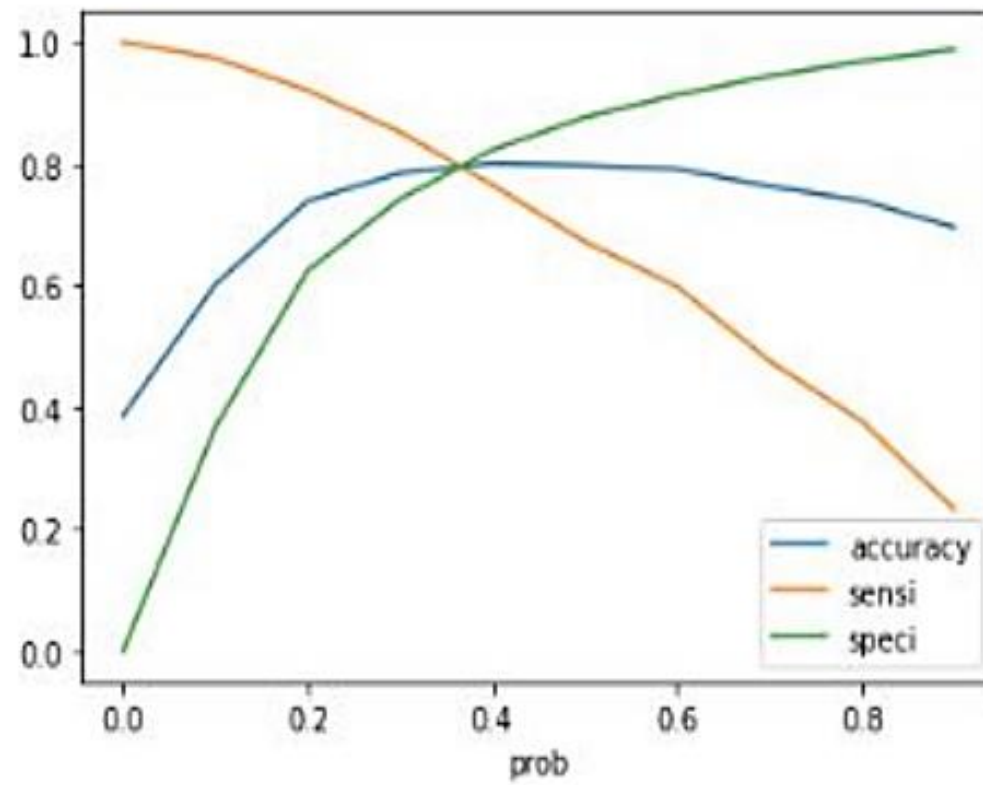
- Splitting the Data into Training and Testing Sets
- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- Use RFE for Feature Selection
- Running RFE with 15 variables as output
- Building Model by removing the variable whose p- value is greater than 0.05 and vif value is greater than 5
- Predictions on test data set
- Overall accuracy 79%

ROC Curve



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

Model Evaluation - Sensitivity and Specificity on Train Data Set

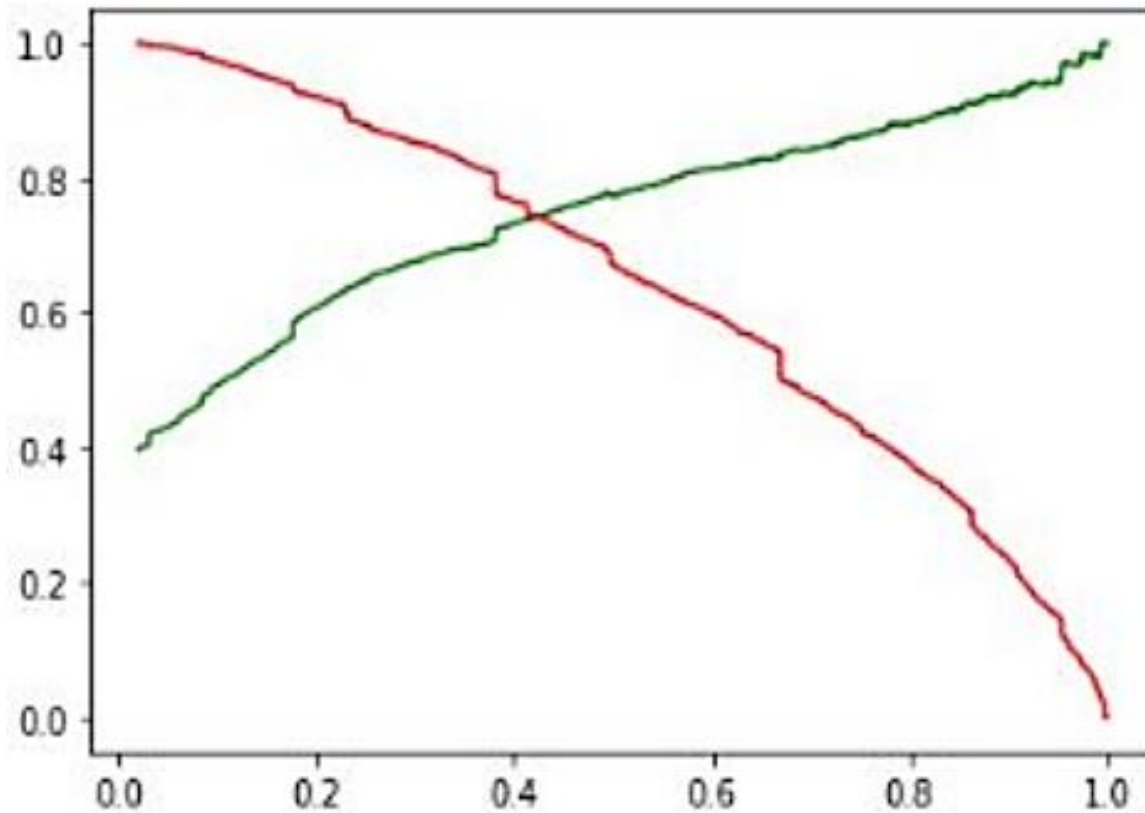


Accuracy=79%

Sensitivity=82%

Specificity=77%

Model Evaluation- Precision and Recall on Train Dataset



The graph depicts an optimal cut off of 0.42 based on Precision and Recall

• Precision - 73 % • Recall - 76%

Model Evaluation – Sensitivity and Specificity on Test Dataset

Overall Accuracy=80%

Precision=71%

Recall=76%

Conclusion

It was found that the variables that mattered the most in the potential buyers are (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
- When the last activity was:
 - a. SMS
 - b. chat conversation
- When the lead origin is Lead add format.
- 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