LEAD SCORING CASE STUDY

by

Harshkumar Vaghmaria

Vaibhav Yadav

Amol Sanjay Shelke

PROBLEM STATEMENT

- Education offers online courses for purchase for business professionals.
- Although education receives a lot of leads, it has an extremely low lead conversion rate.
- For instance, only approximately 25 of 100 leads they could gather in a day might actually be converted.
- The goal of the business is to find the most promising leads, commonly referred to as "Hot Leads," in order to increase the efficiency of this process.
- The lead conversion rate should increase if they are successful in identifying 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 is curious about the most promising leads.
- They aim to create a model that can identify the hot leads for that purpose.
- The deployment of the model for upcoming use.

SOLUTION APPROACH

a) Data cleaning and manipulation.

    1. Verify and deal with duplicate data.

    2. Verify and handle missing and NA values.

    3. Remove columns from the analysis if they have a significant number of missing values.

    4. If required, value impugnation.

    5. Examine and manage data outliers.

b) EDA

    1. Analysis of univariate data: value count, variable distribution, etc.

    2. Bivariate data analysis, including patterns between the variables and correlation coefficients.

c) Scaling Feature & Dummy the data's variables and encoding.

d) Logistic regression is a classification technique that is used to create models and make predictions.

e) Validation of the model.

f) Presentation for Model.

g) Observations and suggestions.

DATA CLEANING AND MANIPULATION

a) 37 total rows and 9240 total columns make up this matrix.

b) Single value features like "Magazine", "Receive More Updates About Our Courses", "Update me on Supply", "Chain Content", "Get updates on DM Content", "I agree to pay the amount through cheque" etc. have been removed.

c) Removing any information from the "Prospect ID" and "Lead Number" that is not required for the analysis.

d) The features that we have dropped include "Do Not Call", "What matters most to you in choosing a course", "Search", "Newspaper Article", "X Education Forums", "Newspaper", "Digital Advertisement", etc. after examining the value counts for some of the object type variables.

e) Dropping the columns with more than 40% of their values missing, such as "Lead Profile" and "How did you hear about X Education".
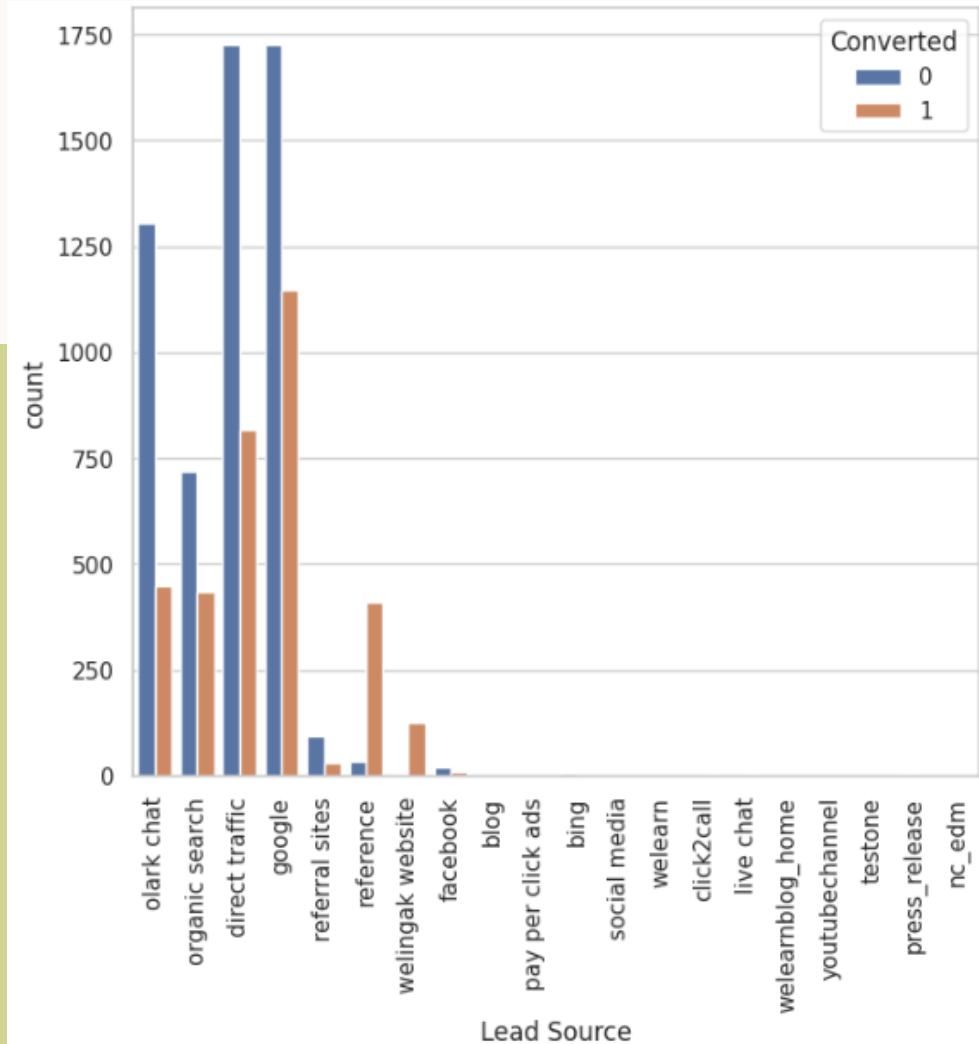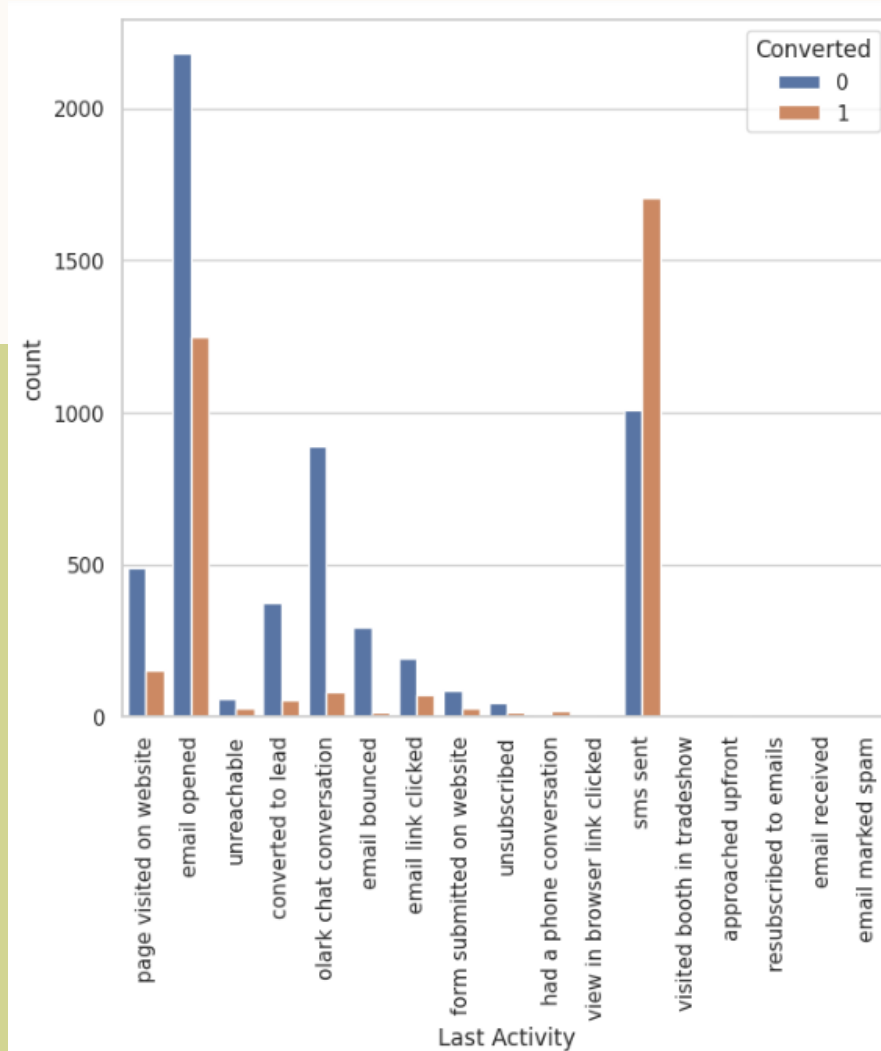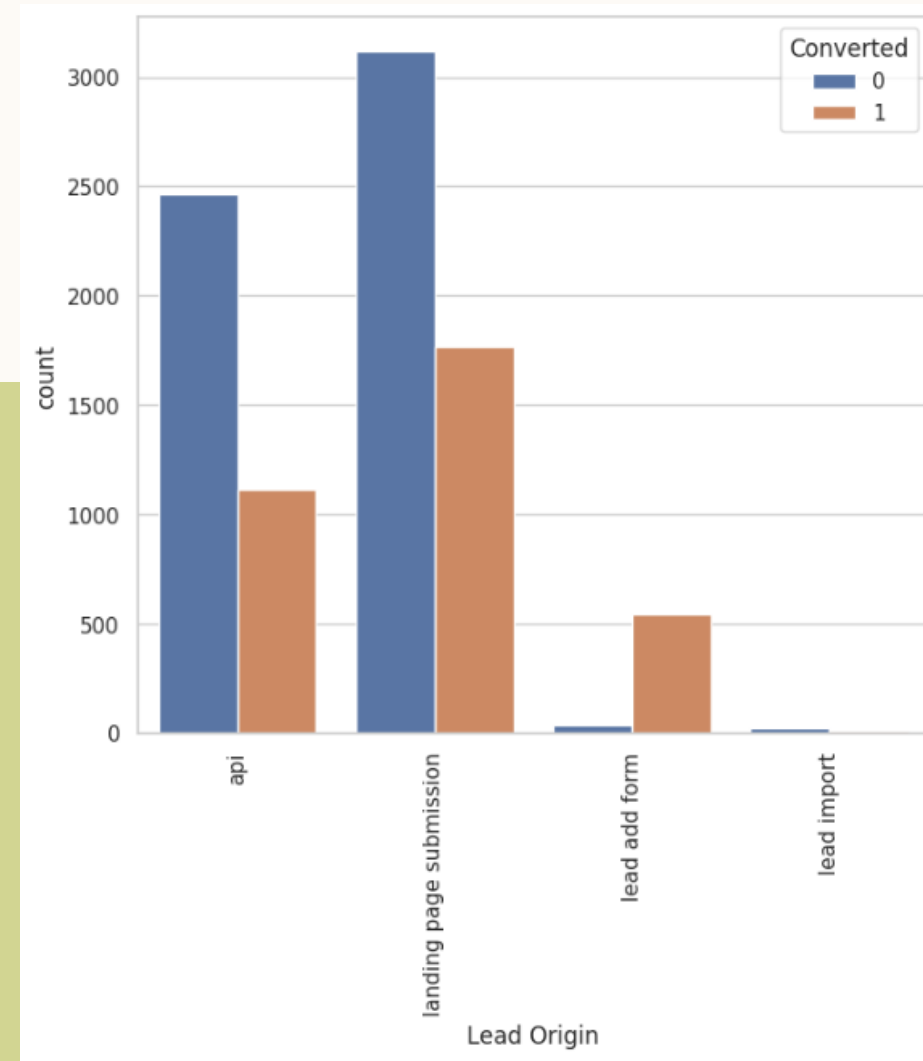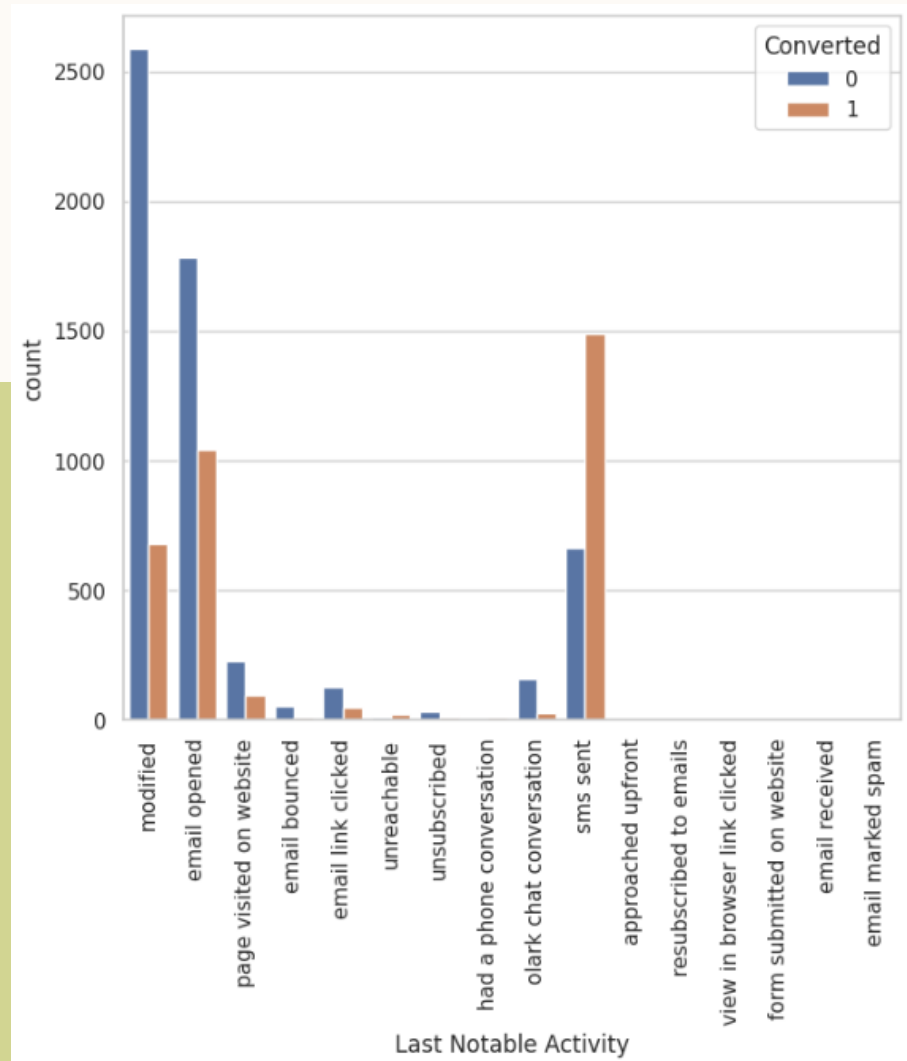
EDA

Catagotical
variables

EDA
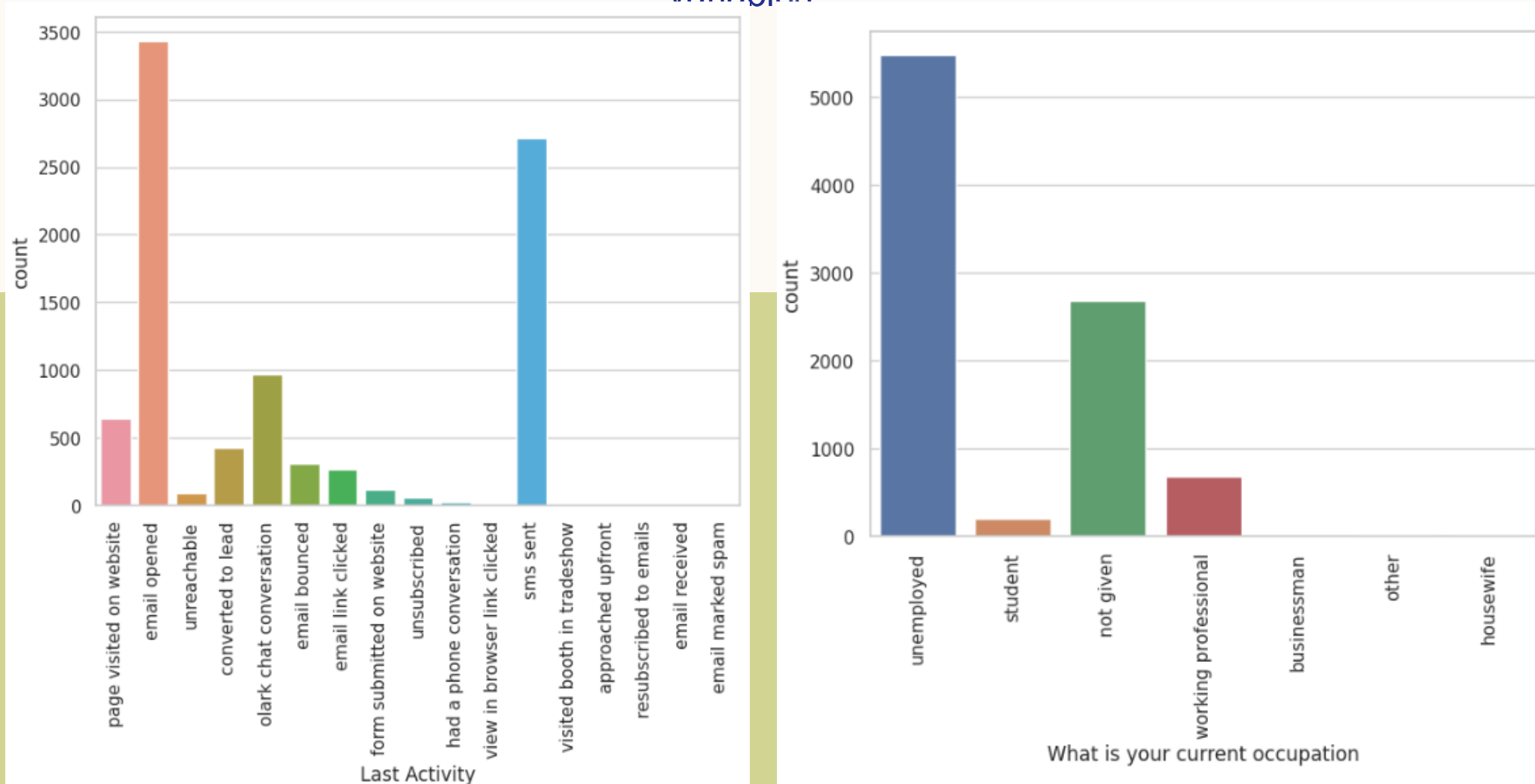
Catagotical variables relation

EDA

Catagotical variables relation

DATA CONVERSION

- For object type variables, dummy variables are constructed.
- Normalization of numerical variables
- Total rows for analysis: 8792.
- Total columns for analysis: 43.

MODEL CREATION

- The first fundamental step in regression analysis is to separate the data into training and testing sets.
- For this split, we used a 70:30 ratio.
- Using RFE for feature selection.
- Running RFE with 15 variables as output.
- Building a model by omitting variables with p-values higher than 0.05 and VIF values higher than 5
- Making predictions using the test data set and achieving an overall accuracy of 81%.
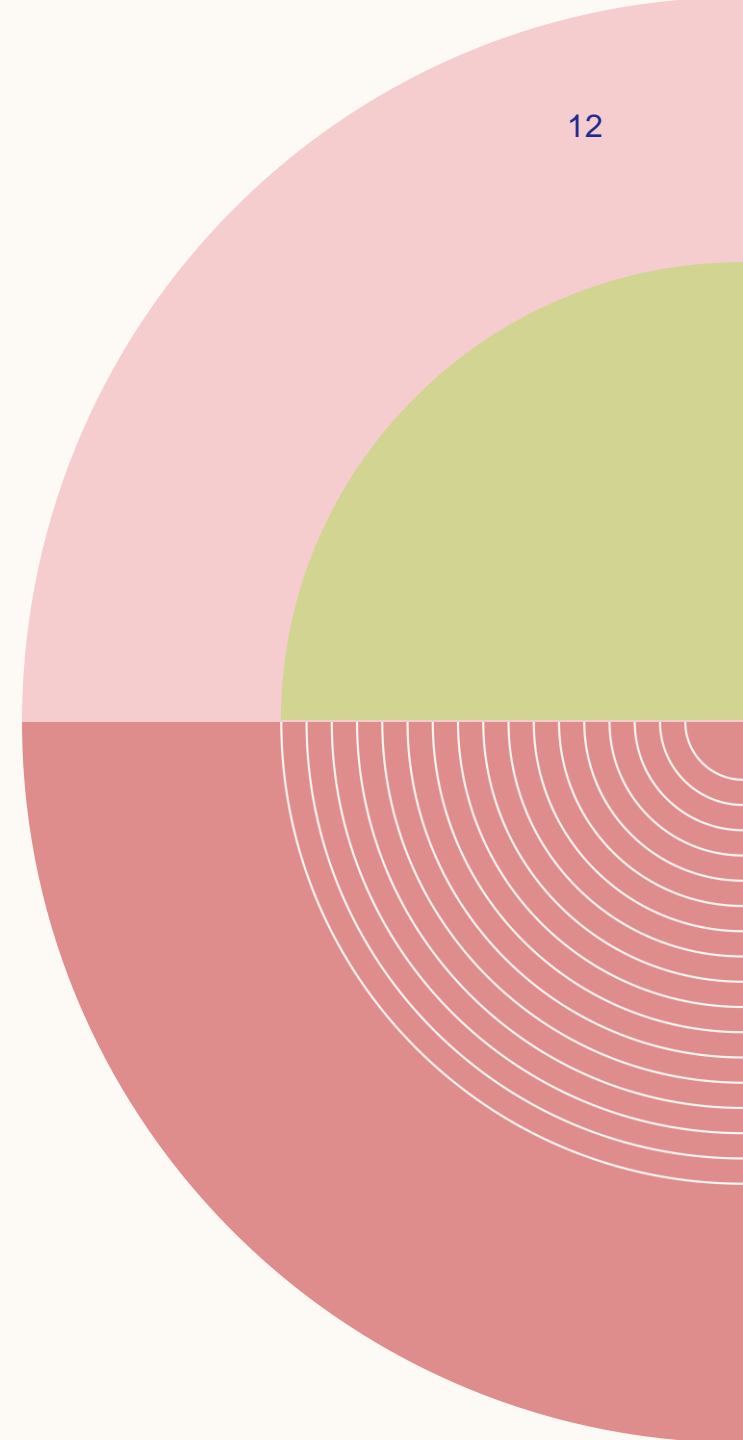
ROC CURVE
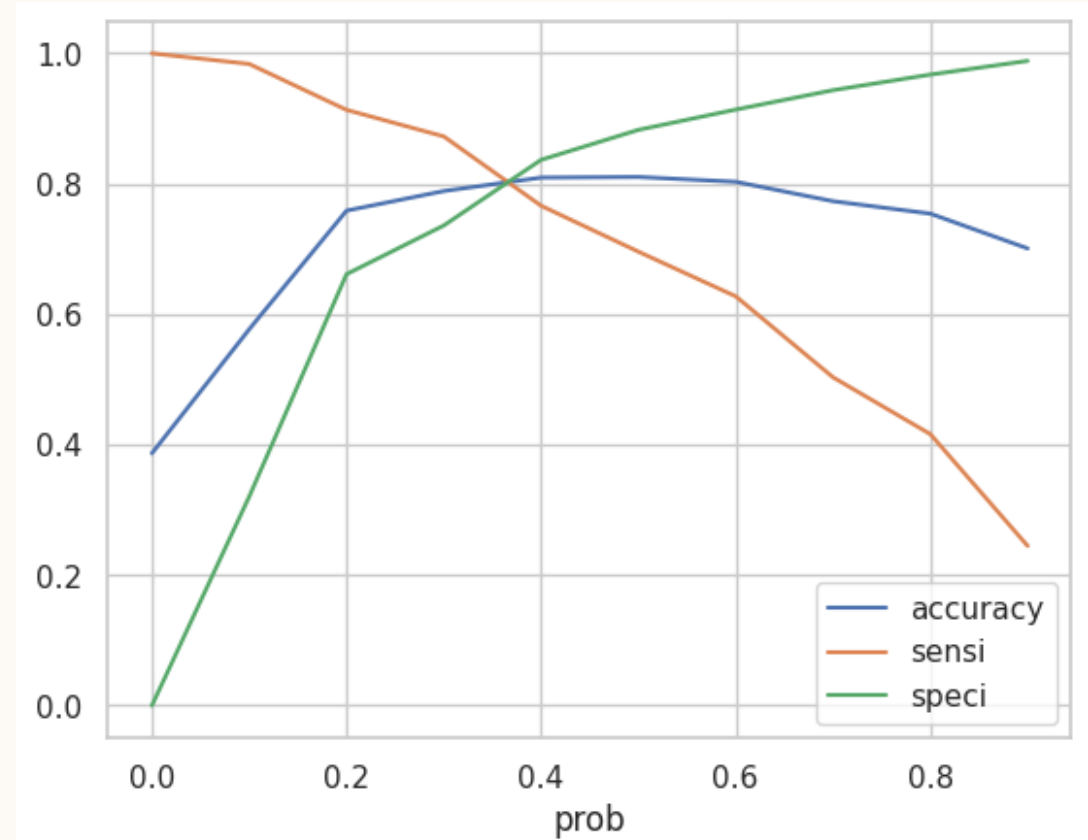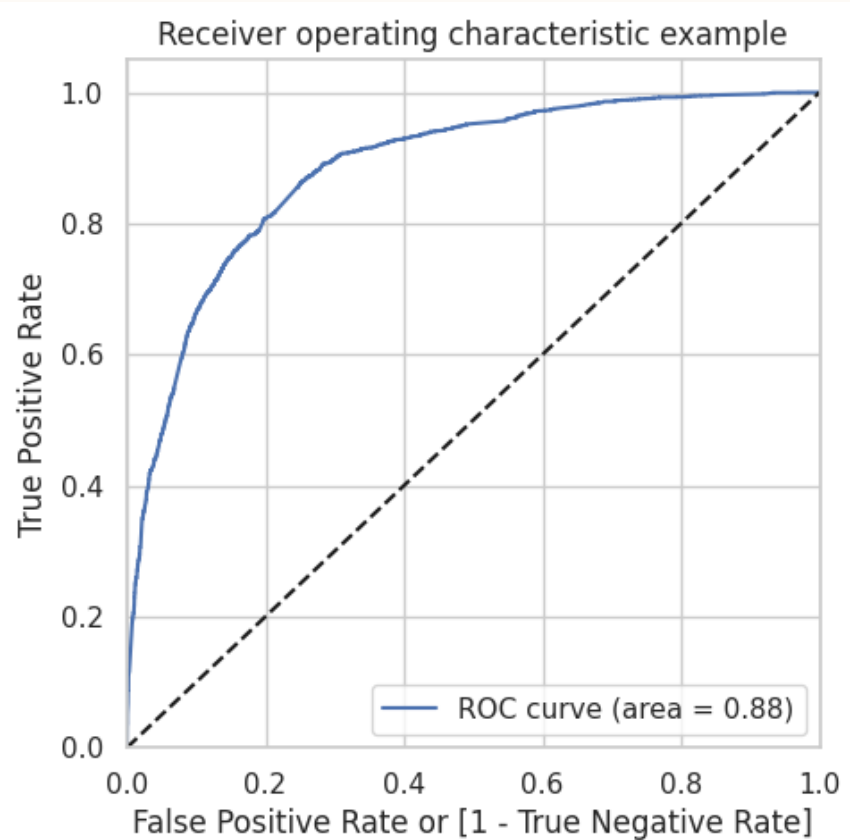
- Determining the Best Cutoff Point.
- The chance where sensitivity and specificity are balanced is the optimal cutoff probability.
- It is clear from the second graph that 0.35 is the ideal cut off.

SUMMARY

- The factors that affected potential purchasers the most were discovered to be (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

4. When the last activity was:

      a) SMS

      b) Olark chat conversation

5. When the lead origin is Lead add format.

6. When the lead is a working professional in their current position. 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.

THANK YOU