

Lead Scoring Case Study

Assignment

GROUP MEMBERS:

1. SRINIVAS DASARI
2. INDRAJEET CHAUDHARY
3. MOHAMMAD ALI

Understanding Business Problem

About the Business & Its Problem

- ▶ X Education company sells online courses to industry professionals
- ▶ Markets its courses through online channels
- ▶ Online form fill up is through these channels is a lead for the company
- ▶ **Problem statement** is company gets lot many leads but conversion rate is low, i.e. 30%

Business Objective

- ▶ Objective for increasing the lead conversion rate
- ▶ Finding out **hot leads** more promising for conversion
- ▶ Building a model predicting most promising hot leads using past data
- ▶ Deploying model for future use

Solution Approach

Data Manipulation & Analysis

- ▶ Understanding data - working on past data
- ▶ Data cleaning
- ▶ EDA – Exploratory Data Analysis
- ▶ Dummy Variables Creation & Features scaling
- ▶ Data Preparation for Model building and evaluation – Test Train split

Model Building & Evaluation

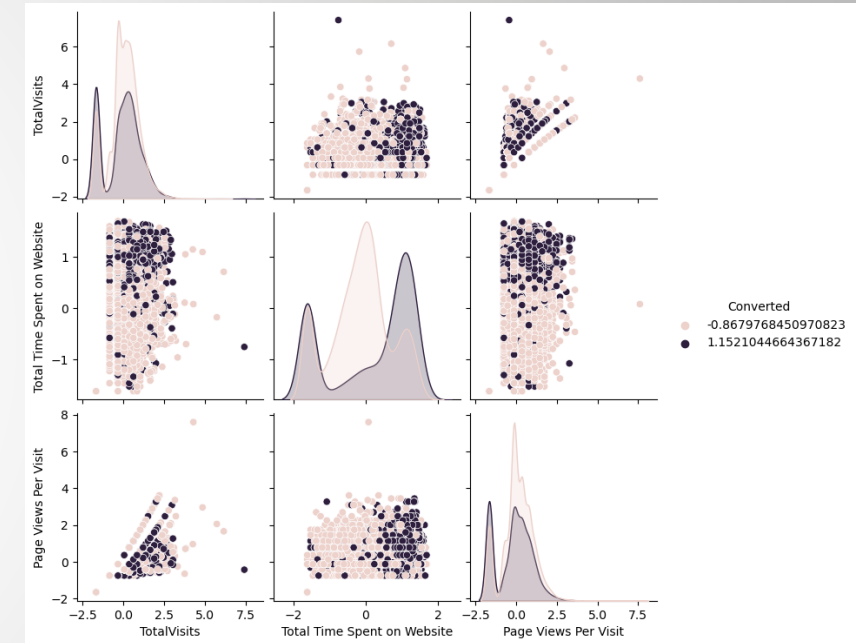
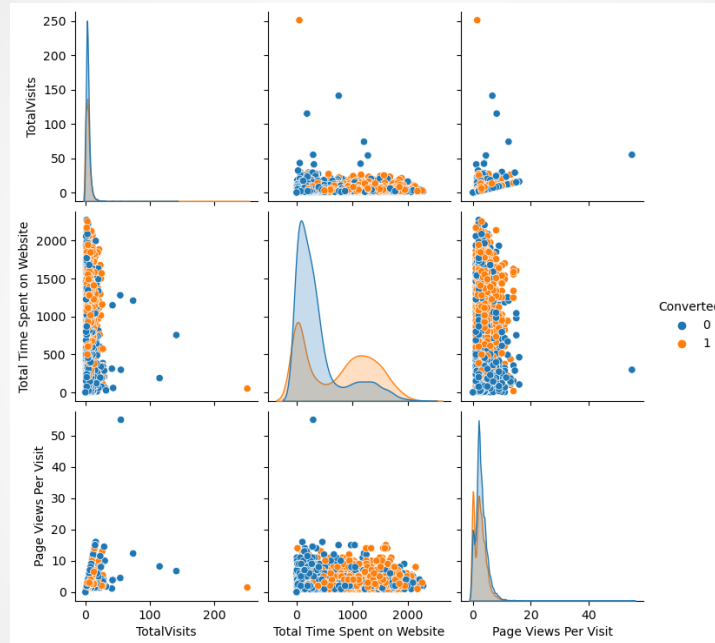
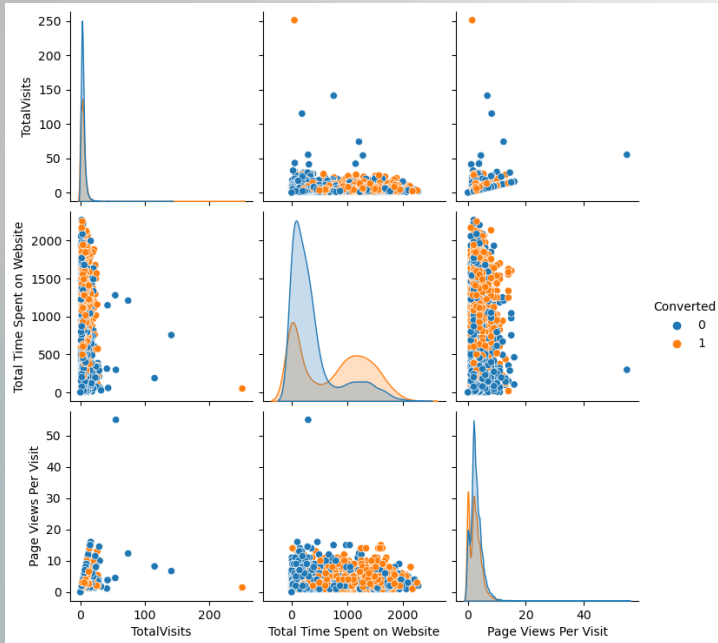
- ▶ Using Logistic Regression RFE method for model building
- ▶ Feature selection using P value and VIF
- ▶ Model evaluation and Prediction using Accuracy, Sensitivity, Specificity and ROC curve
- ▶ Model evaluation and Prediction using Precision and Recall

Data Understanding & Manipulation

- ▶ Total Number of Rows =37, Total Number of Columns =9240.
- ▶ Removing the variables which are not necessary for the analysis
- ▶ Dropping the columns having more than 3000 missing values
- ▶ Dropping the columns which have values as 'Select' and not found useful for analysis
- ▶ Dropping the rows based wherever null values are there in few variables

Data Preparation for Modelling

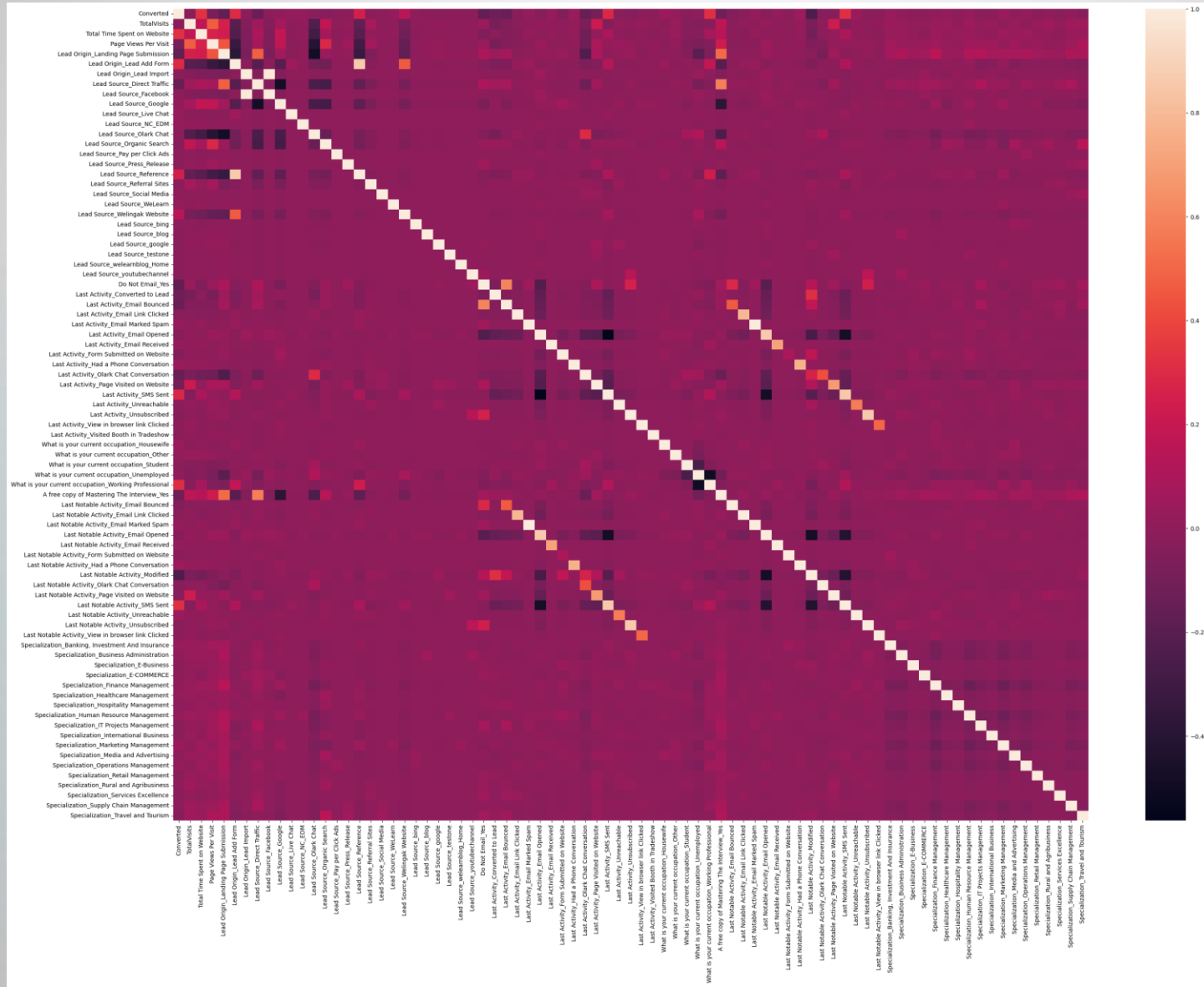
Numerical Variables Analysis



Data Conversion

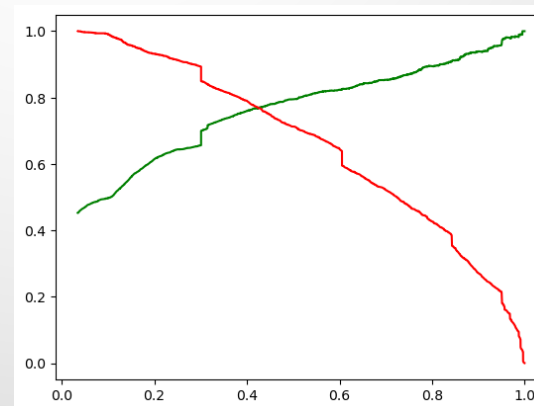
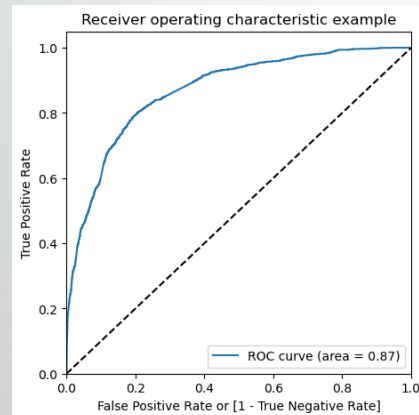
- ▶ Dummy Variables are assigned to Categorical variables, deleted the duplicates columns
- ▶ Numerical variables are normalised
- ▶ Total Rows for Analysis: 7643, Total columns for Analysis: 81 – post dummy variable creation

Correlation among variables



Model Building & Evaluation

- ▶ Splitting the Data into Training and Testing Sets into 70:30 ratio
- ▶ Using 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
- ▶ Prediction values calculated using cutoff 0.44 derived from ROC curve and Precision Recall Curve
- ▶ Predictions on test data set
- ▶ Overall accuracy was 80%, Precision 75% and Recall 76%



Summary

Recommended Actions for Improvement in Leads Conversion

- ▶ Need to nurture the potential leads well by way of educating the leads about the product and constantly engaging in order to get a higher lead conversion
- ▶ Sort out the best prospects from the generated leads which contribute most towards the probability of a lead getting converted - below are 3 main dependent variables
 - ▶ 'TotalVisits', 'Total Time Spent on Website', 'Page Views Per Visit'
- ▶ Leads may be informed about new courses, services, job offers and future higher studies
- ▶ Monitor each lead carefully so that tailored information can be shared with them
- ▶ Job offerings, information on courses that suits best according to the interest of the leads may be shared
- ▶ A proper plan to chart the needs of each lead will go a long way to capture the leads as prospects
- ▶ Focusing on converted leads by way of holding question-answer sessions with leads to extract the right information that made them opting for course
- ▶ Making further inquiries and appointments with the leads to determine their intention and mentality to join online courses