LEADS SCORING CASE STUDY

DS C57

Valkal Bairagi

Business Objective

01. Leads

▶ X education wants to know most promising leads

02. Model

▶ For that they want to build a Model which identifies the hot leads.

03. Future Ready

▶ Deployment of the model for the future use

Introduction

▶ 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 the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Methodology

Data cleaning and data manipulation

- 1. Check and handle duplicate data.
- 2. Check and handle NA values and missing values.
- 3. Drop columns, if it contains a large number of missing
- values and are not useful for the analysis.
- 4. Imputation of the values, if necessary.
- 5. Check and handle outliers in data.
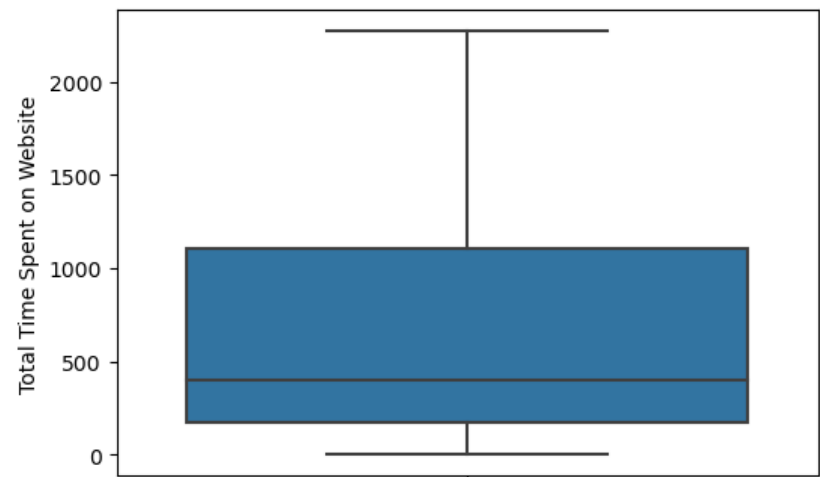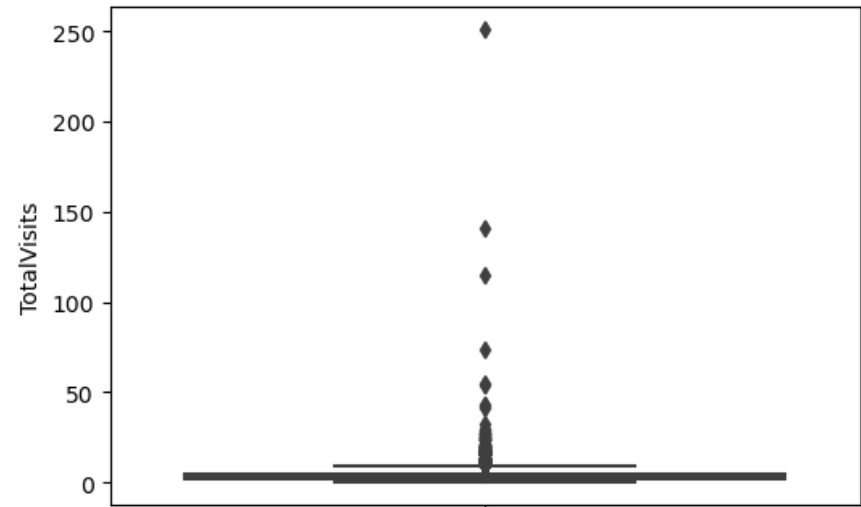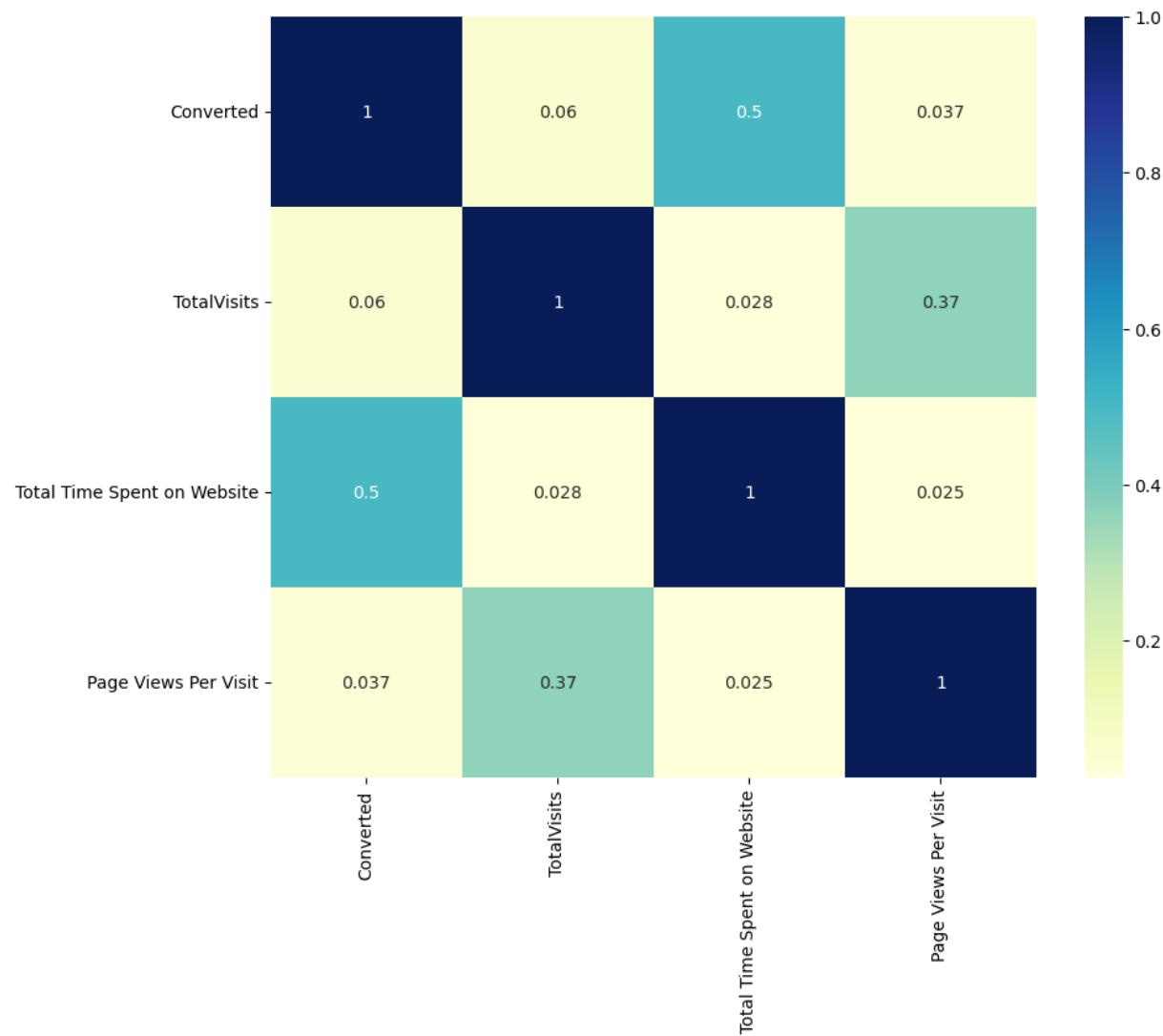
Exploratory Data Analysis (EDA)

- 1. Univariate data analysis: value count, distribution of variables, etc.
- 2. Bivariate data analysis: correlation coefficients and pattern between the
- variables etc.
- 3. Feature Scaling & Dummy variables and encoding of the data.
- 4. Classification technique: logistic regression is used for model making
- and prediction.
- 5. Validation of the model.
- 6. Model presentation.
- 7. Conclusions and recommendations.

Annual Review December 19, 2023

Data Manipulation

▶ Total Number of Rows=37,Total Number of Columns =9240.

▶ Single value features like "Magazine" , "ReceiveMoreUpdates About Our Courses" , "Update my supply"

▶ Chain Content" , "Get updates on DM Content" , "I agree to pay the amount through cheque" etc. have been dropped.

▶ Removing the "ProspectID" and "Lead Number" which are not necessary for the analysis.

▶ After checking for the value counts for some of the object type variables, we find some of the features which have enough variance, which has dropped, the features are: "Do Not Call" , "What matters most to you in choosing course" , "Search" , "Newspaper, Article" , "XEducation Forums" , "Newspaper" , "DigitalAdvertisement" etc.

▶ Dropping the column shaving more than 35% as missing values such as 'How did you hear about X Education' and ' Lead Profile'.

5

Numerical Attribute Analysis
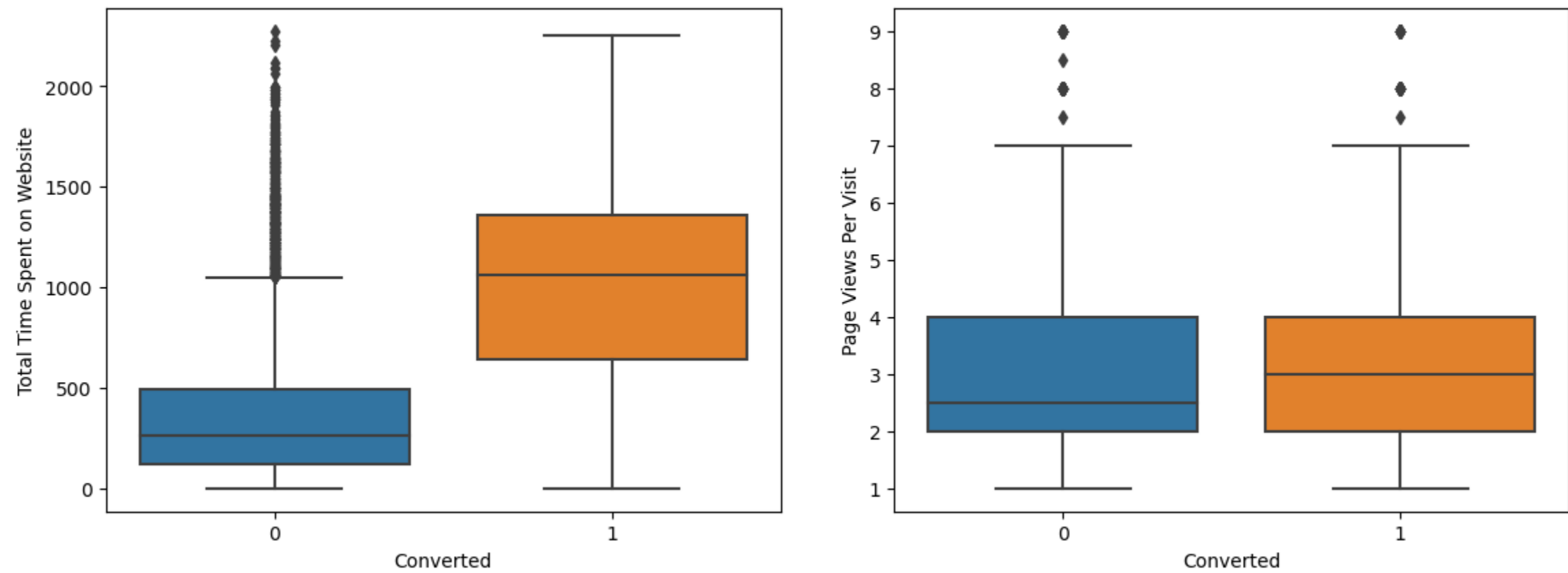
Data Conversion

▶Numerical Variables are normalized

▶ Dummy Variables are created for object type variables

▶ Total Rows for Analysis: 9240

▶ Total Columns for Analysis: 37

Model Building

1

▶ 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.

3

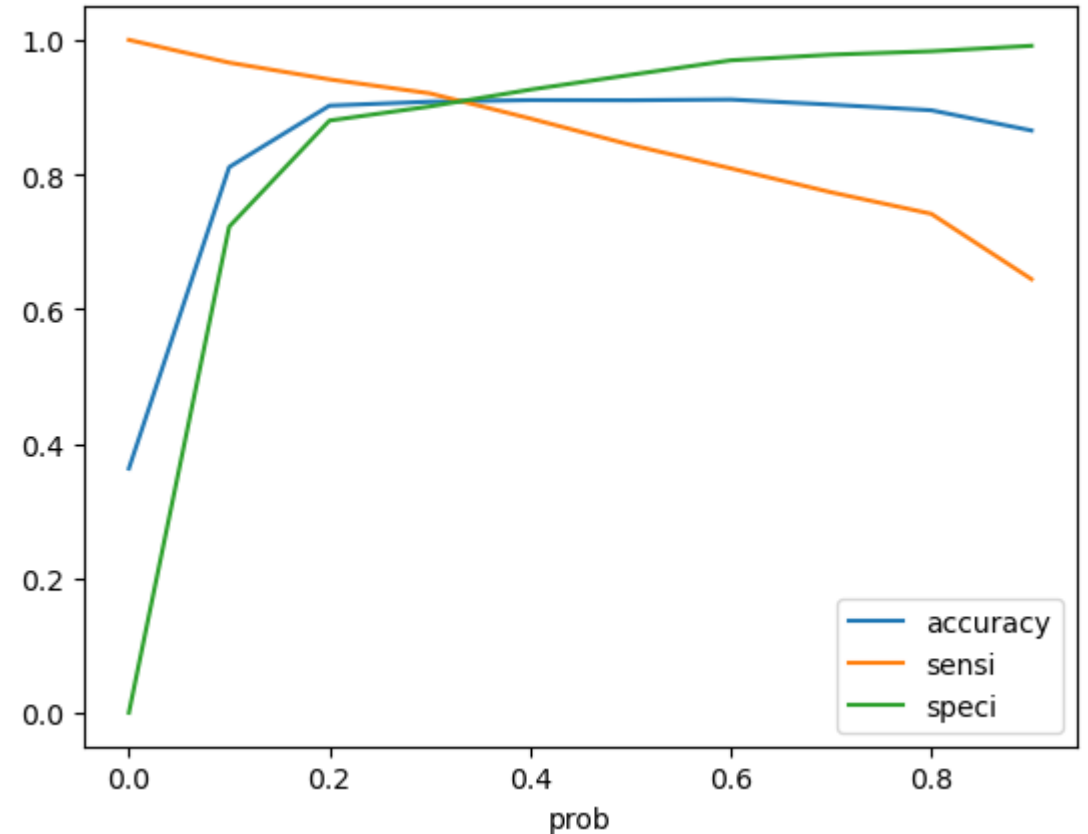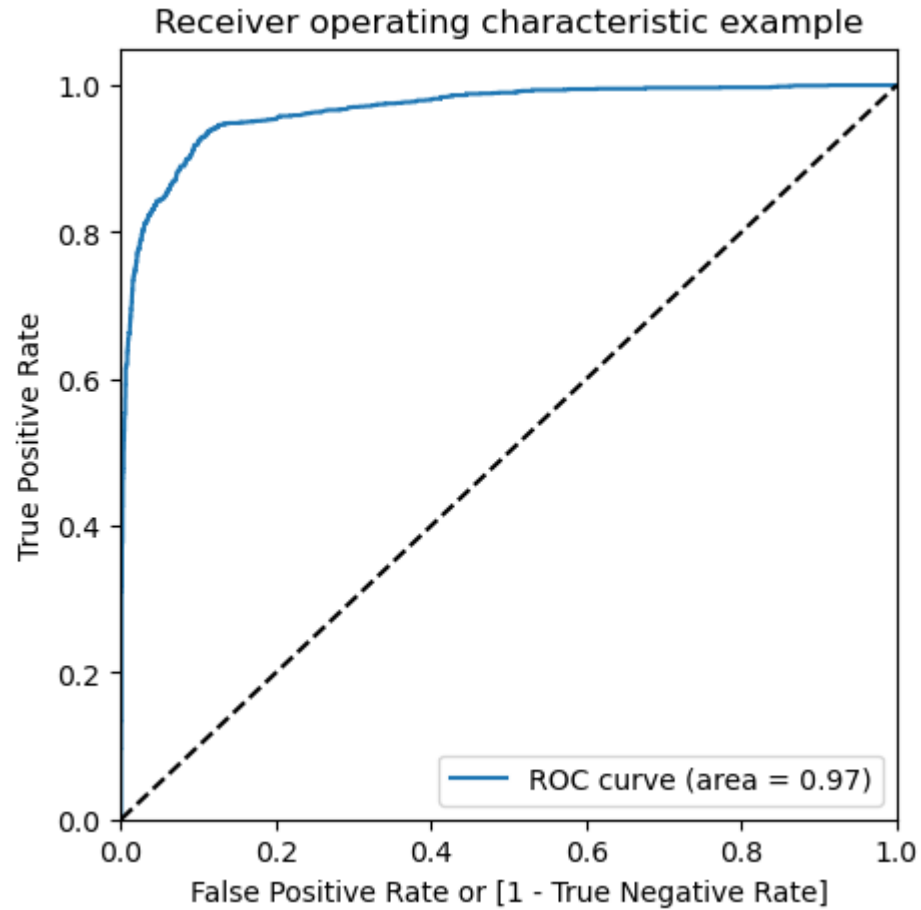▶ Building Model by removing the variable whose p-value is greater than 0.05 and vi value is greater than 5

2

▶ Use RFE for Feature Selection

▶ Running RFE with 15 variables as output

4

▶ Predictions on test data set

▶ Overall accuracy 81%

Annual Review   December 19, 2023

ROC Curves

The ROC Curve should be a value close to 1. We are getting a good value of 0.97 indicating a good predictive model.

PREDICTION ON TEST SET

1

- ▶ Before predicting on the test set, we need to standardize the test set and need to have exact same columns present in our final train dataset.

- ▶ After doing the above step, we started predicting the test set, and the new prediction values were saved in a new data frame.

- ▶ After this we did model evaluation i.e. finding the accuracy, precision, and recall.

2

- After running the model on the Test Data these are the figures we obtain:

- Accuracy: 90.92%

- Sensitivity : 91.41%

- Specificity: 90.62%

3

- The values obtained for Train & Test:

Train Data:

- Accuracy: 90.81%

- Sensitivity: 92.05%

- Specificity: 90.10%

Test Data:

- Accuracy: 90.92%

- Sensitivity: 91.41%

- Specificity: 90.62%

Summary

Our business is good

The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model

Annual Review December 19, 2023