LEAD SCORING CASE STUDY

Focused Business Approach Using Logistic Regression Technique

LEADS SCORING CASE STADY

By Chetram Mairha, Chandrakant Sangam & Chitwan Tayal

Problem Statements

- 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.
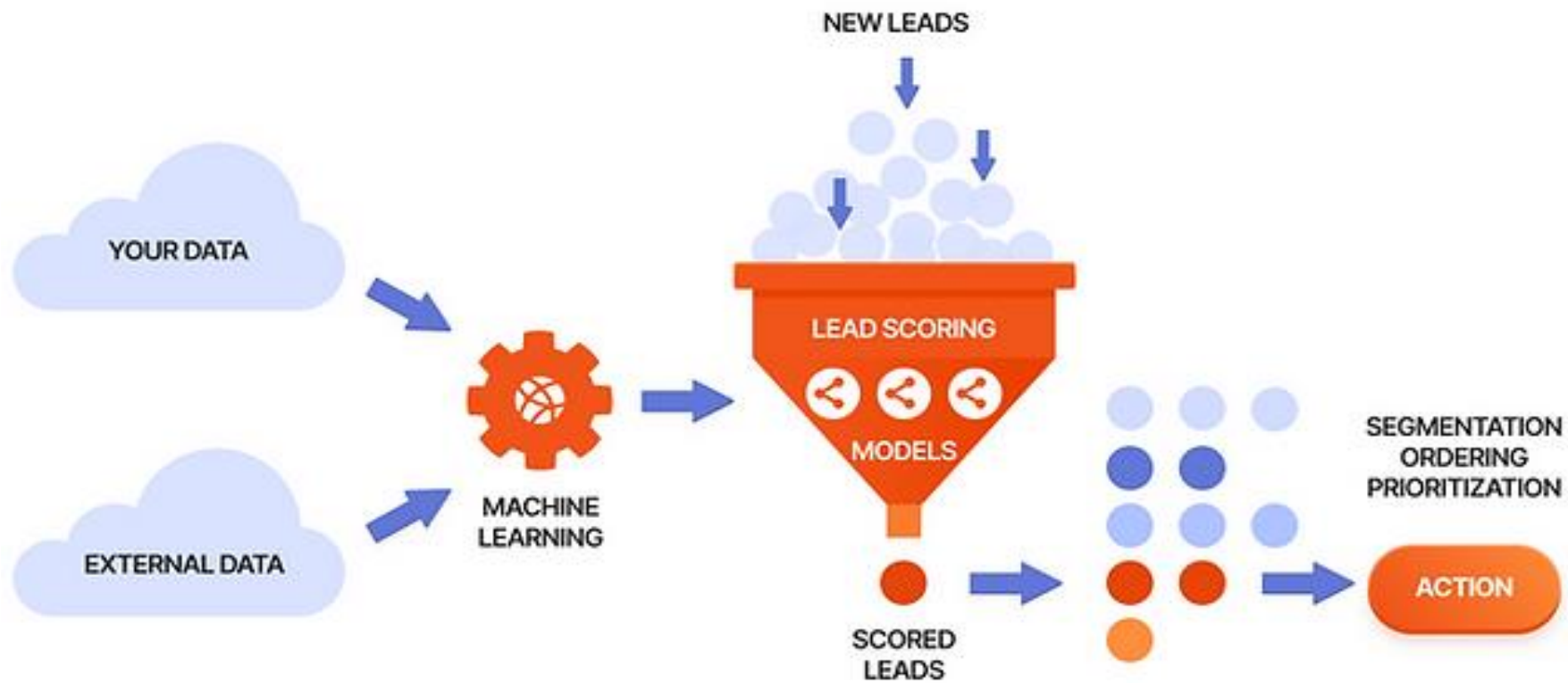
BUSINESS OBJECTIVE

- 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

SOLUTION 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. Classification technique: logistic regression is used for model making and prediction.

4. Validation of the model.

5. Model presentation.

6. Conclusions and recommendations.

DATA MANIPULATION

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

▶ Single value features like "Magazine" , "Receive More Updates About Our Courses" , "Update my supply"

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

▶ After checking for the value counts for some of the object type variables, we find some of the features which have enough variance, which have dropped, the features are: "Do Not Call" , "What matters most to you in choosing course" , "Search" , "Newspaper, Article" , "X Education Forums" , "Newspaper" , "Digital Advertisement" etc.
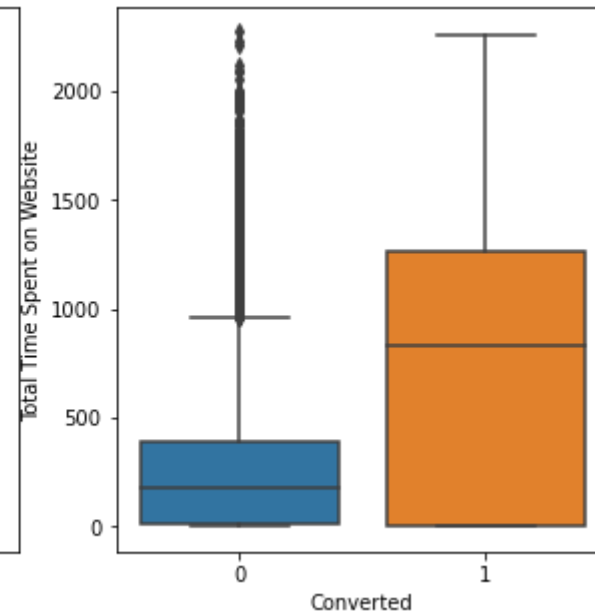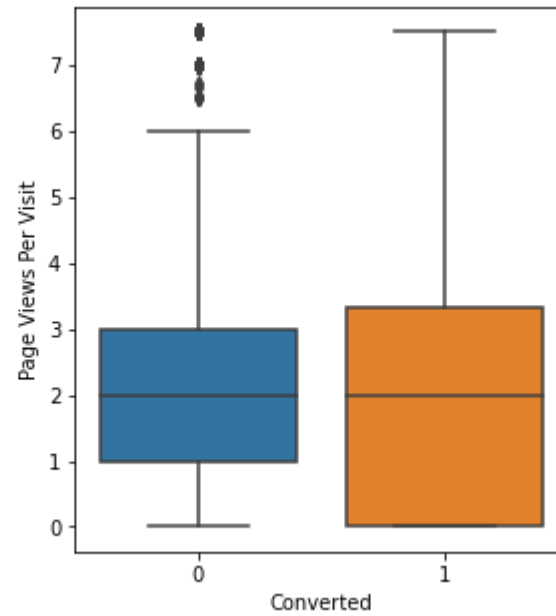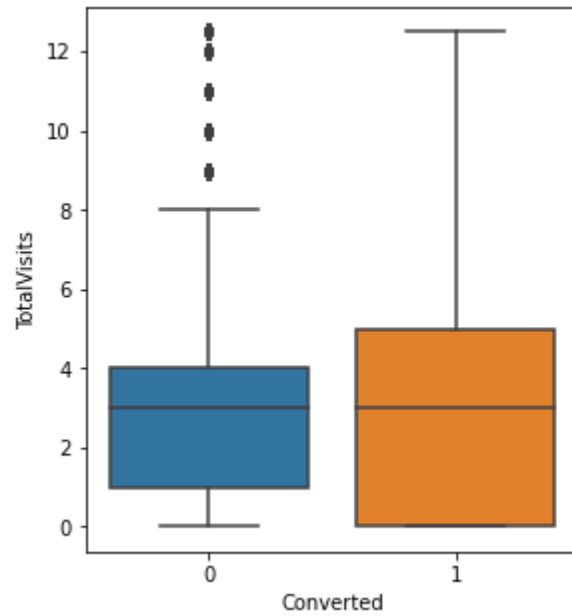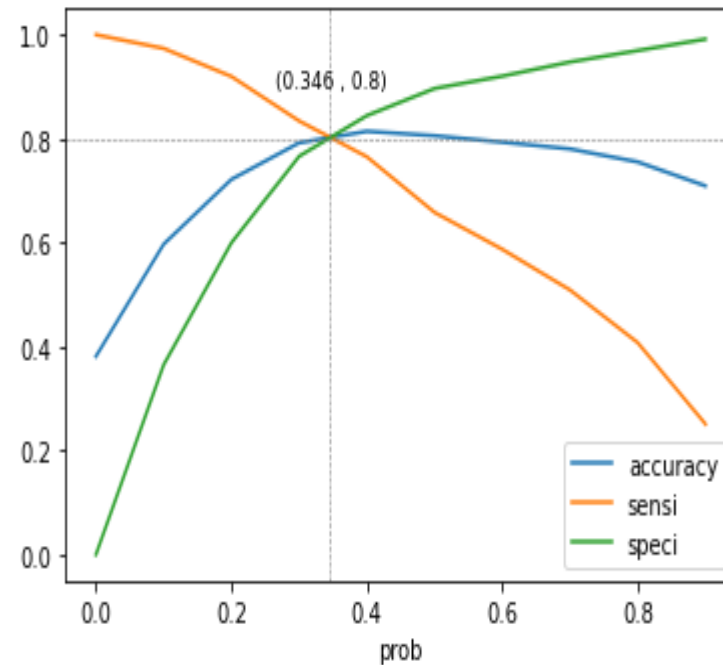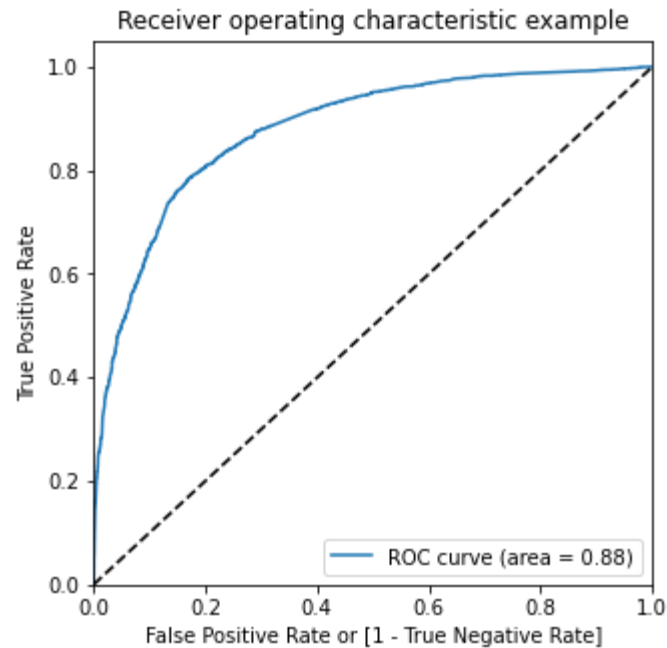
EXPLORATORY DATA ANALYSIS (EDA)
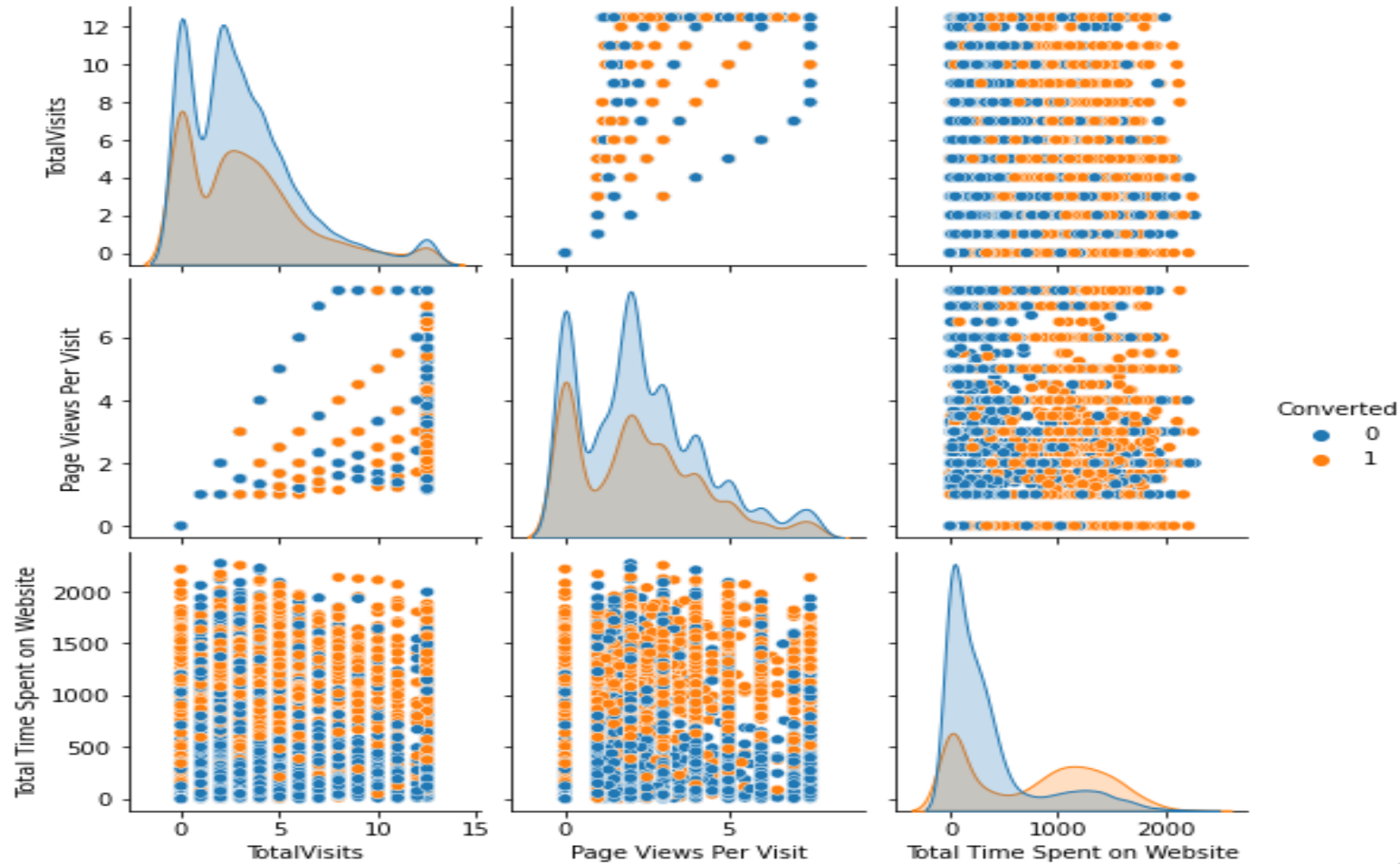
Heat Map
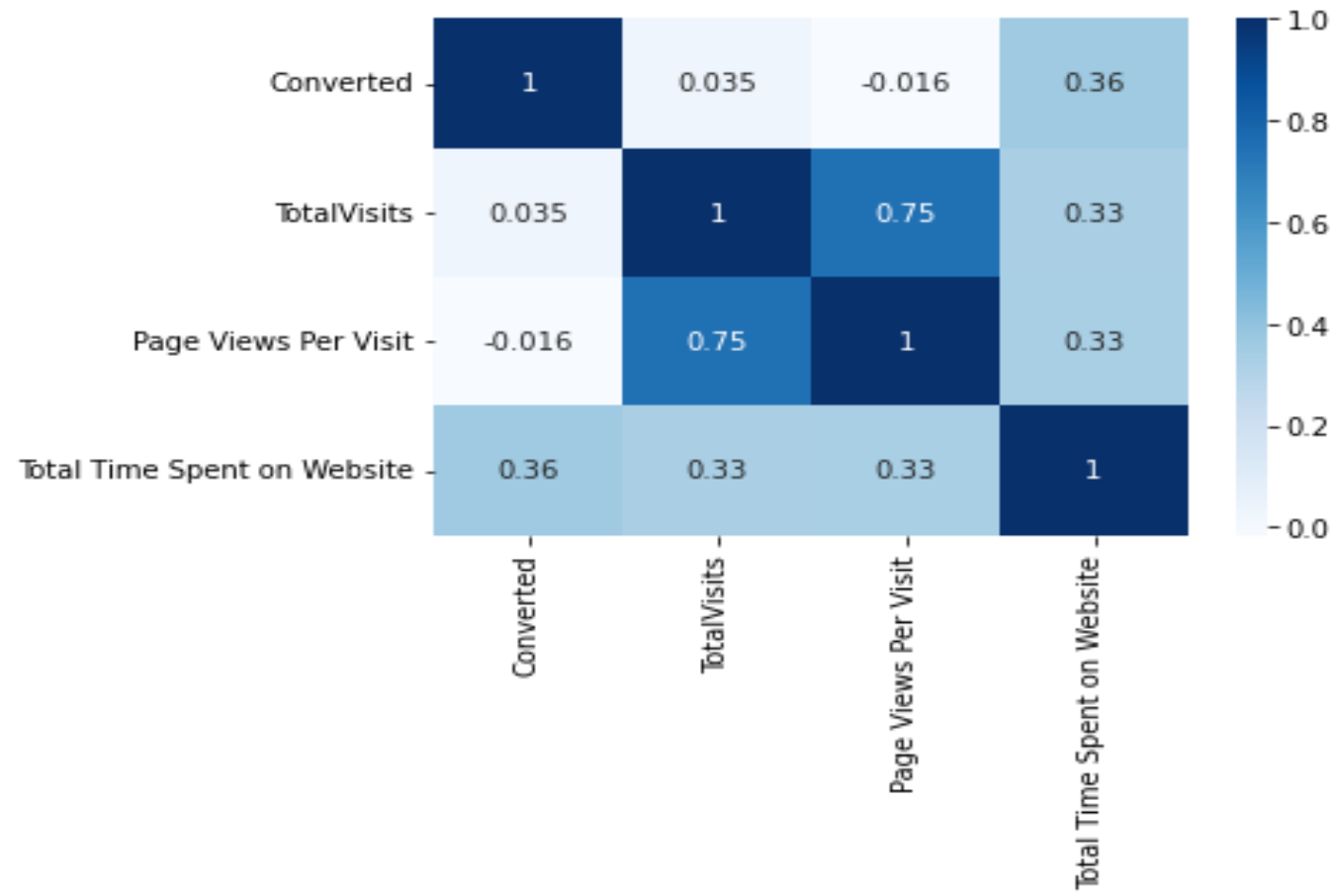
BOX PLOT

MODEL BUILDING

▶ Splitting the Data into Training and Testing Sets

▶ Running RFE with 15 variables as output

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

▶ Predictions on test data set

▶ Overall accuracy 81%

ROC Model

1. Finding Optimal Cut off Point. 2. Optimal cut-off probability is that
3. Probability where we get balanced sensitivity and specificity.
4. From the second graph it is visible that the optimal cut off is at 0.35.

PREDICTION ON TEST SET

- 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.

- This shows that our test prediction is having accuracy, precision, and recall scores in an acceptable range.

- This also shows that our model is stable with good accuracy and recall/sensitivity.

- Lead score is created on test dataset to identify hot leads – high the lead score higher the chance of conversion, low the lead score lower the chance of getting converted

Result of the Model

▶ **Train - Test**

1. Train Data Set:

      Accuracy: 80.46%

      Sensitivity: 80.05%

      Specificity: 80.71%

2. Test Data Set:

      Accuracy: 80.34%

      Sensitivity: 79.82% ≈ 80%

      Specificity: 80.68%

Recommendation To Business

- Top three variables to focus for the probability of a lead gettng converted.
    1. Lead Source_Welingak Website
    2. Lead Source_Reference
    3. Current_occupaƟon_Working Professional

- We could recruit interns and ask them to do follow up over the phone for the Hot leads as listed above.

THANK YOU