LEAD SCORE CASE STUDY

Group Members:

1.Ganesh Killi

2.Shruti Srivastava

PROBLEM STATEMENT:

- X Education sells online courses to industry professionals
- X Education gets a lot of leads, its lead conversation rate is very poor.
- To make this process more efficient, the company wishes to identify the most potential leads also knowns 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 Preparation

  1.Check and handle duplicate data

  2.Check and handle NA values and missing values

  3.Drop columns, if it contains large amount of missing values and 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 variable etc

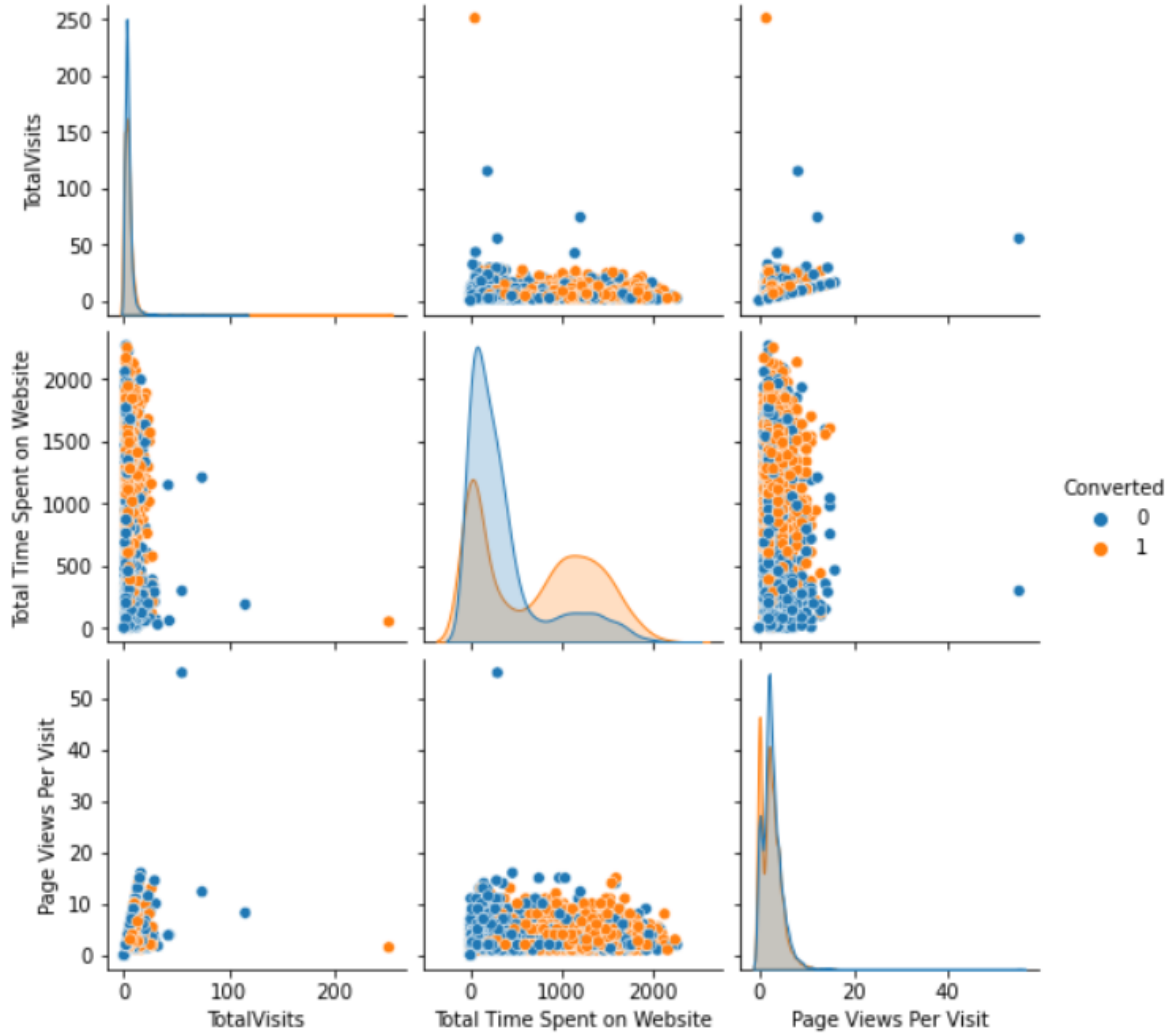
  2. Bivariate Data Analysis : Correlation and pattern between the variables etc

- Preparing the Data for Modelling
- Feature Scaling , Dummy variables and encoding of the data.
- Model Evaluation and Finding the Optimal cut off
- Classification Technique: Logistic regression used for the model making and prediction
- Validation of the model
- Conclusions and recommendations

DATA CLEANING AND PREPARATION:

- ✓ Check the number of missing values in each column
- ✓ Drop all the columns in which greater than 3000 missing values are present
- ✓ Check the number or Percentage of null values in each column
- ✓ Get the value counts of all the columns
- ✓ Drop the columns which has the value SELECT as it would be of no use for the analysis
- ✓ Repeat the above steps for the rows as well

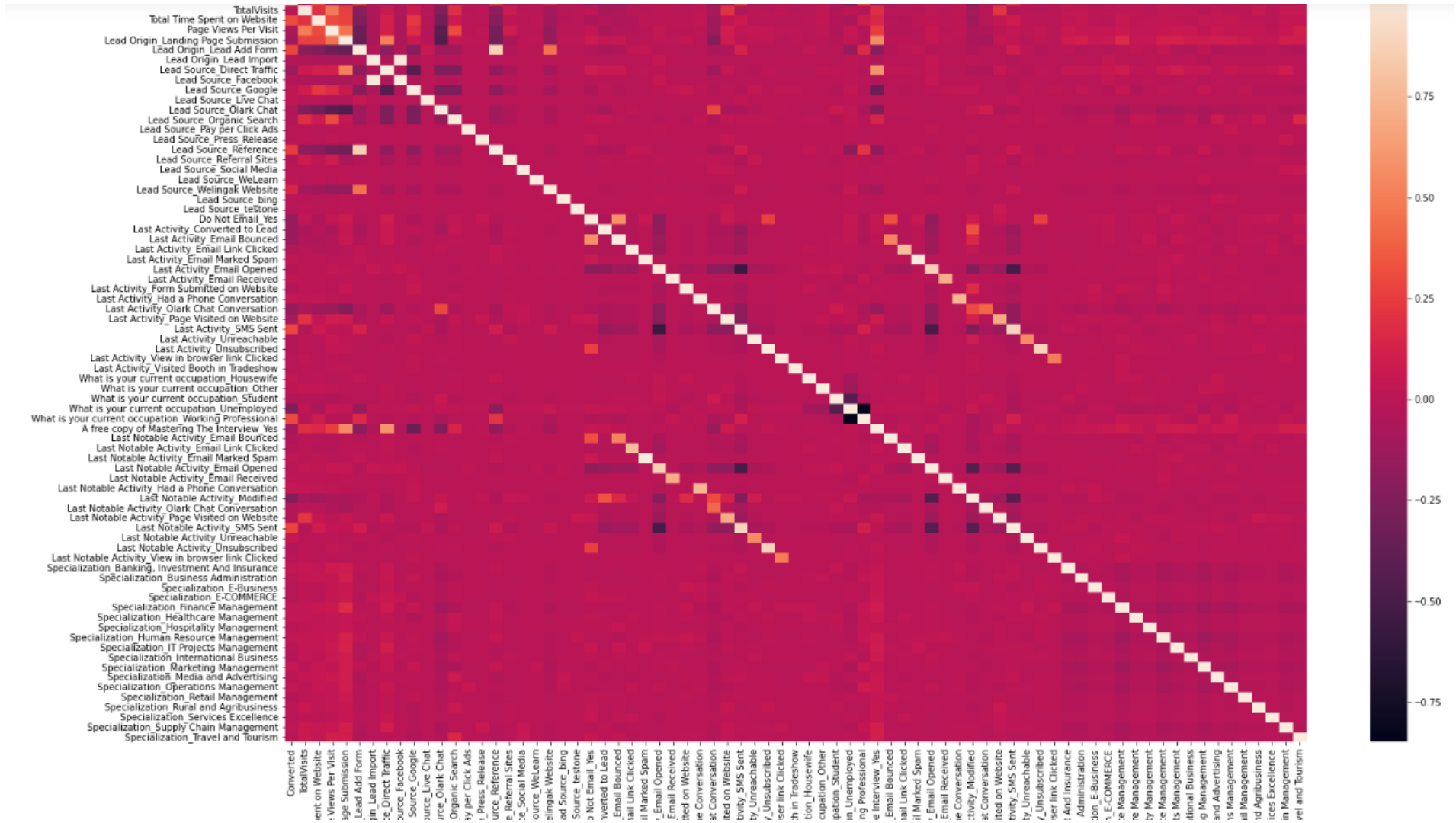
PREPARING THE DATA FOR MODELLING



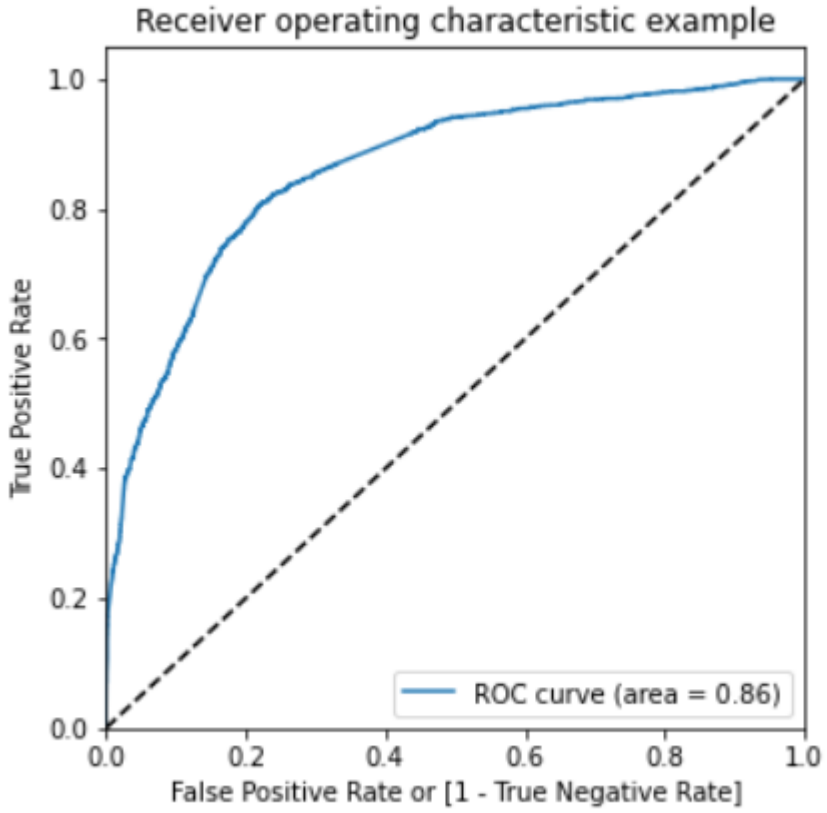
After the data cleaning, below are the 3 columns that can be used for data modelling:

1.Total Visits
2.Total time spent on website
3.Page Views per visit

CORRELATION:

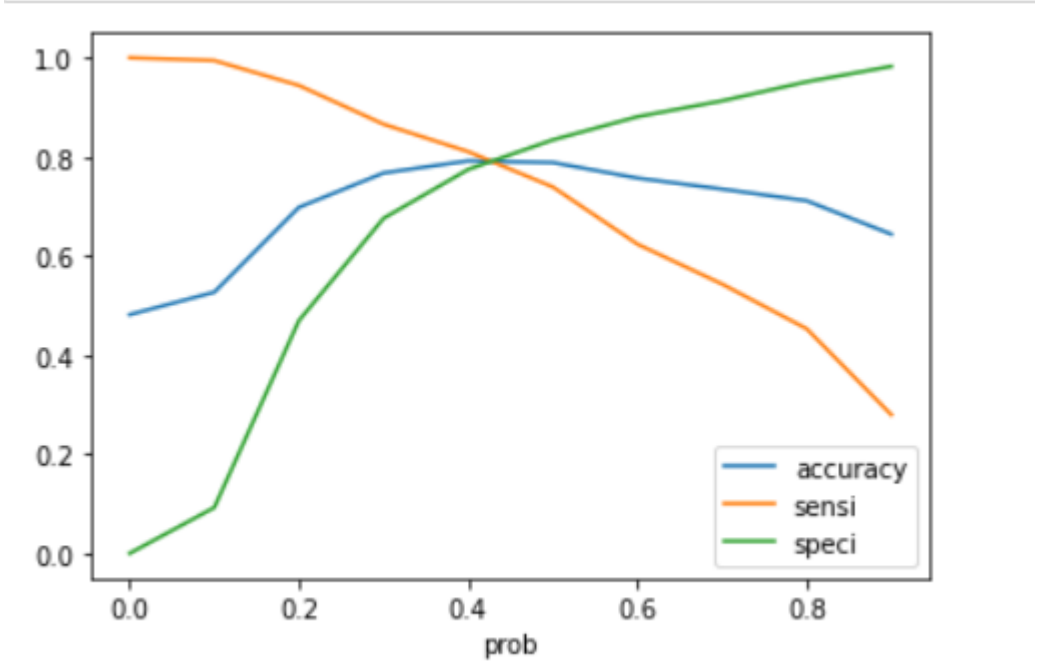


ROC CURVE :



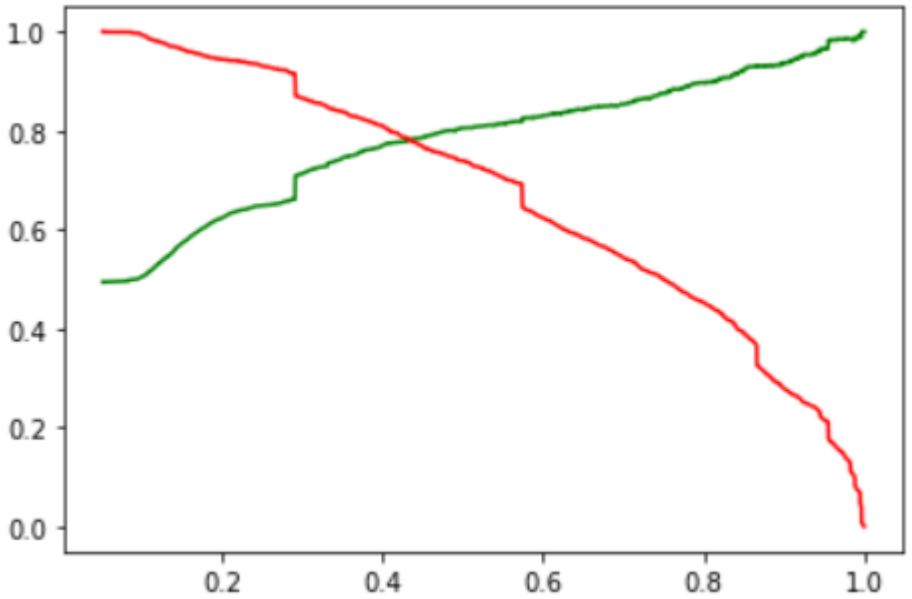
The area under the curve of the ROC is 0.86 which is quite good. So we seem to have a good model. Let's also check the sensitivity and specificity tradeoff to find the optimal cutoff point

ACCURACY, SPECIFICITY AND SENSITIVITY :



As you can see that around 0.42, you get the optimal values of the three metrics. So let's choose 0.42 as our cutoff now.

PRECISION RECALL CURVE:



PREDICTION	TRAIN	TEST
PRECISION	0.78	0.78
RECALL	0.77	0.77

CONCLUSION:

- The lead and customer funnel is the right approach to check the outcome.
- As we can see, there were a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom.
- In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.
- For the instance of checking the conversion rate, firstly sort out the best prospects from the leads you have generated. 'TotalVisits' , 'Total Time Spent on Website' , 'Page Views Per Visit' which contribute most towards the probability of a lead getting converted.
- Secondly, we must keep a list of leads handy so that you can inform them about new courses, services, job offers and future higher studies. Monitor each lead carefully so that you can tailor the information you send to them. Carefully provide job offerings, information or courses that suits best according to the interest of the leads.
- A proper plan to chart the needs of each lead will go a long way to capture the leads as prospects. Focus on converted leads. Hold question-answer sessions with leads to extract the right information you need about them.
- Make further inquiries and appointments with the leads to determine their intention and mentality to join online courses.