LEAD SCORING CASE STUDY

Focused business approach using logistic regression technique

Group Members:

Abhishek Ranjan
Alisha Kadam
Parushni Singh

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

To help X Education select most promising leads (Hot Leads), i.e. the leads that are most likely to convert into paying customers.

Selection of hot leads

Focused marketing

Higher lead conversion rate
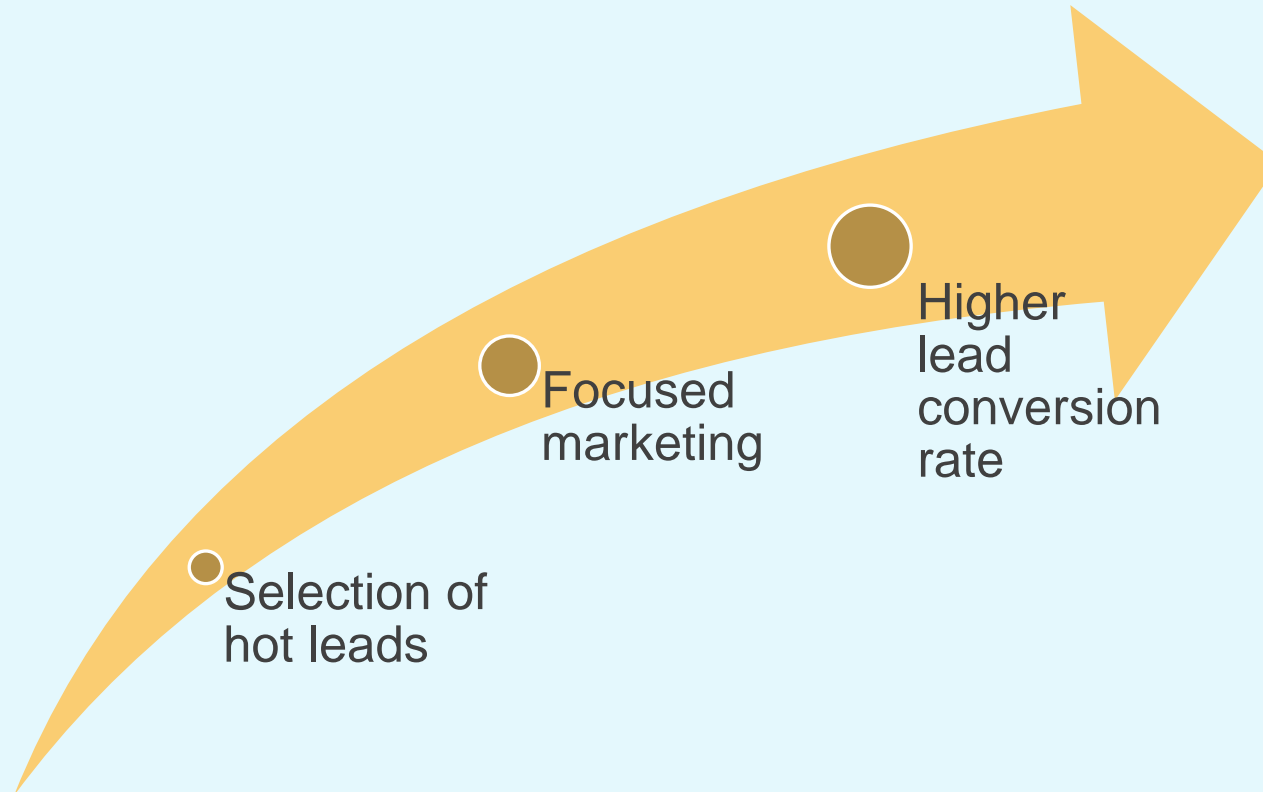
Methodology

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

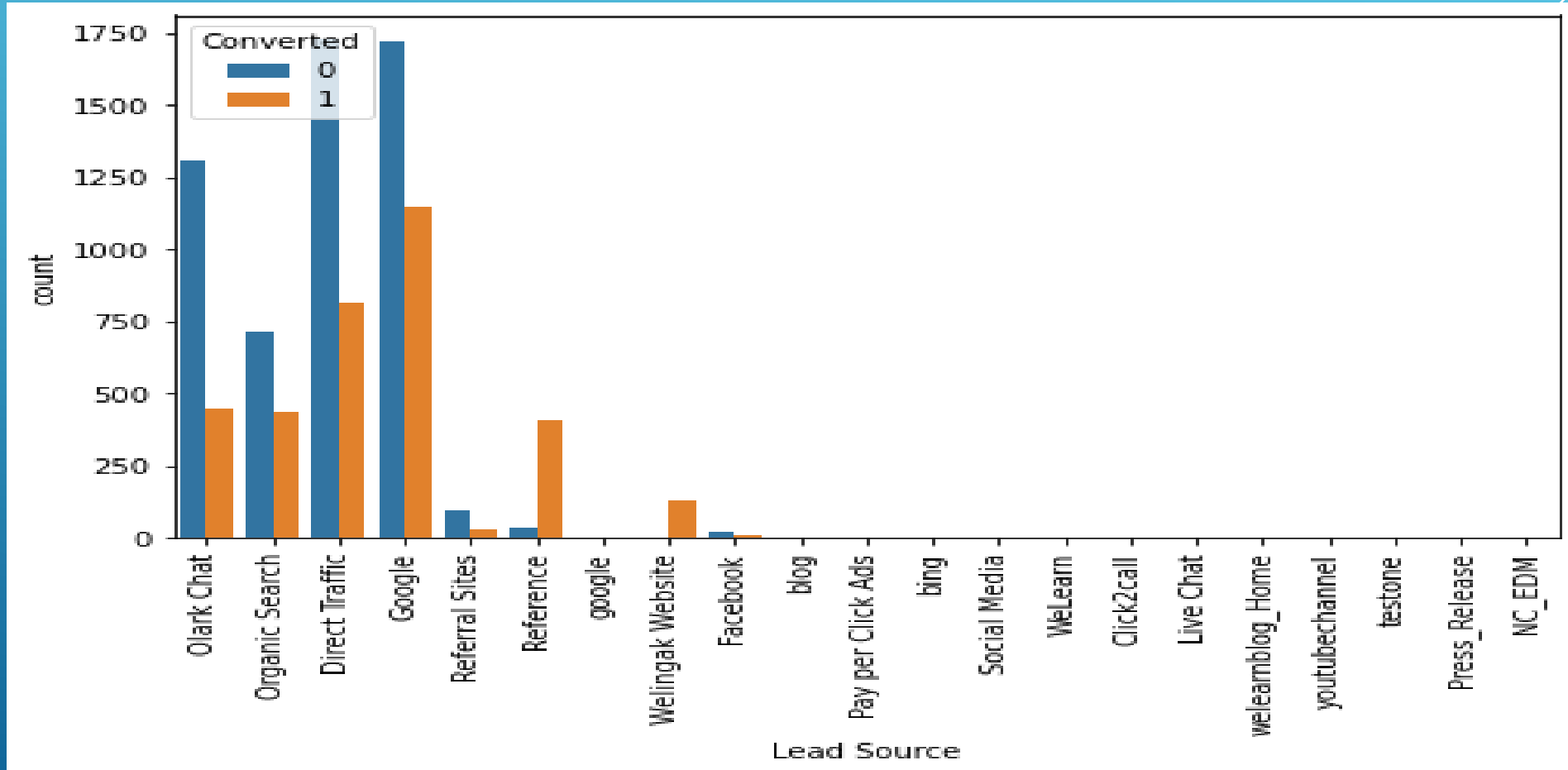**Preparing the Data for Modelling**

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

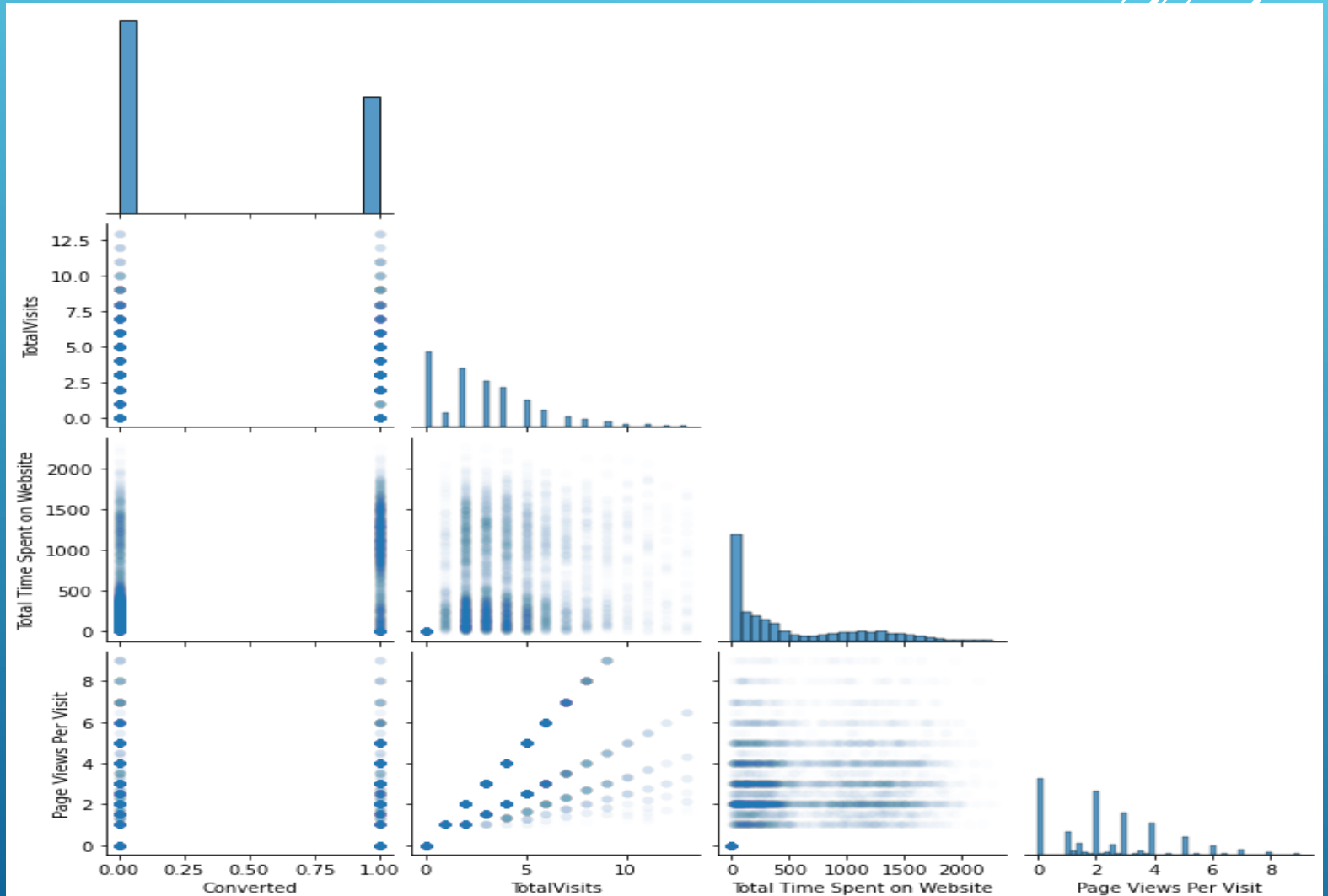➢ Repeat the above steps for the rows as well

Lead Origin
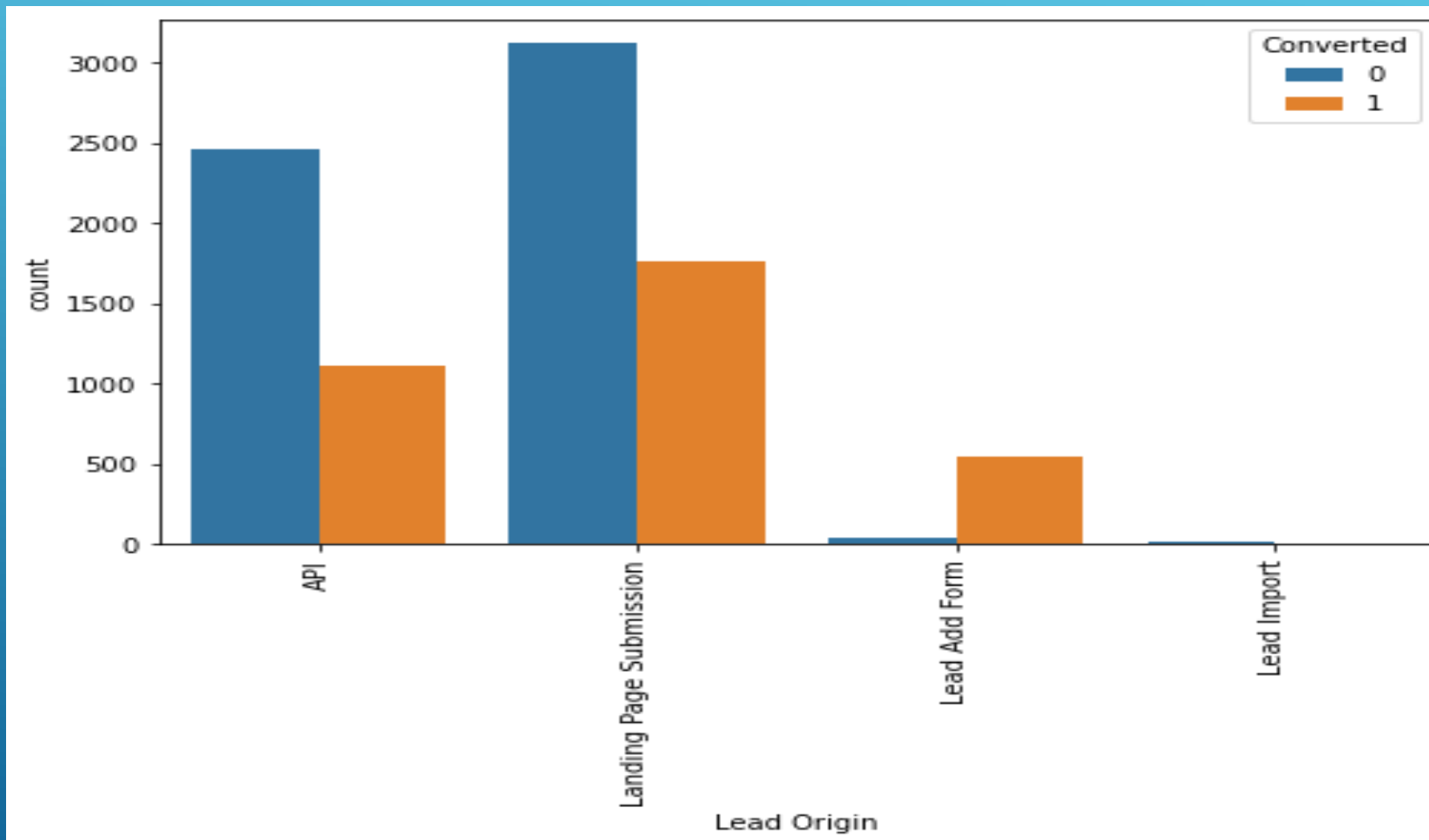
Lead Source

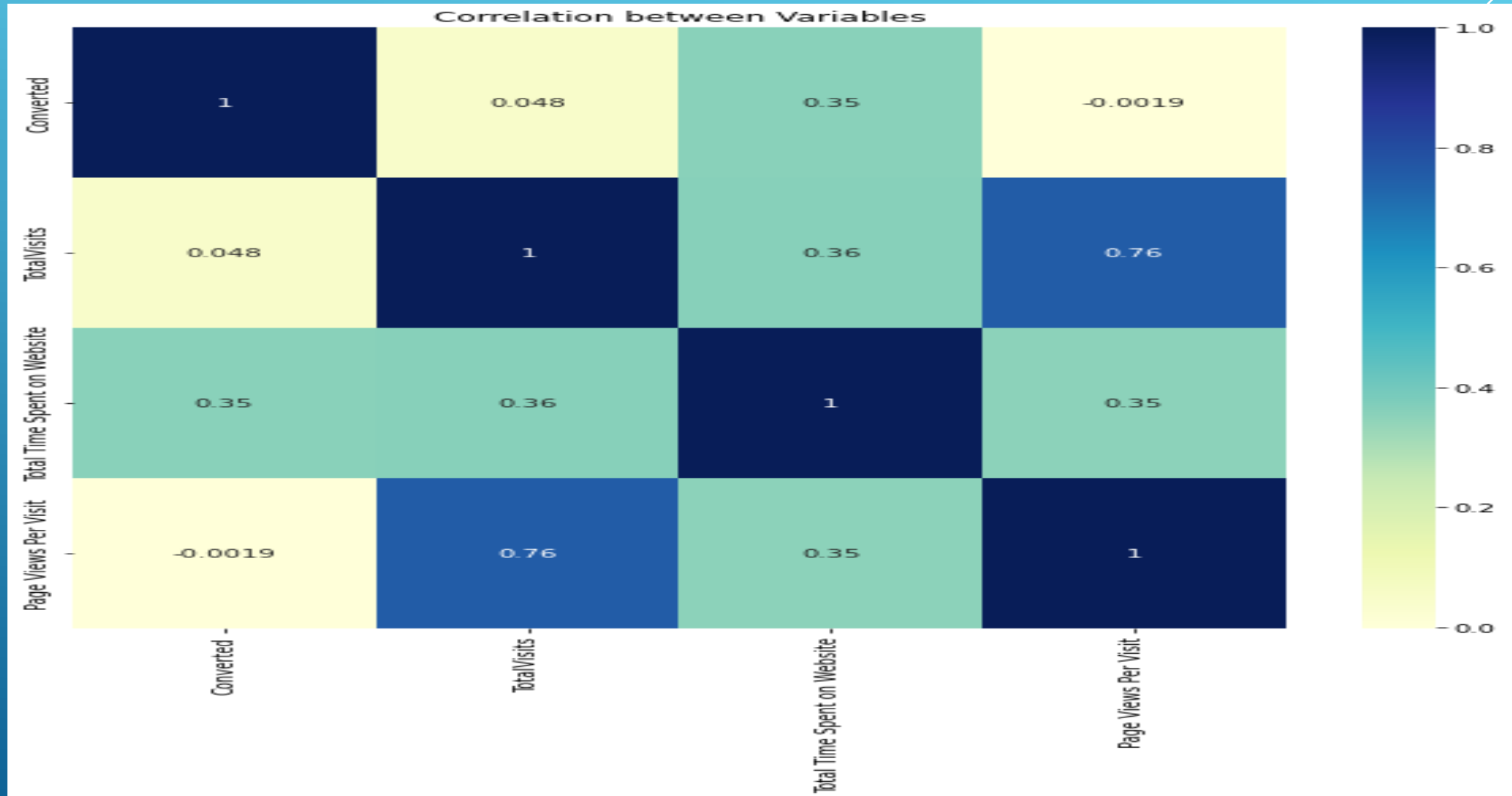# Preparing The Data For Modelling

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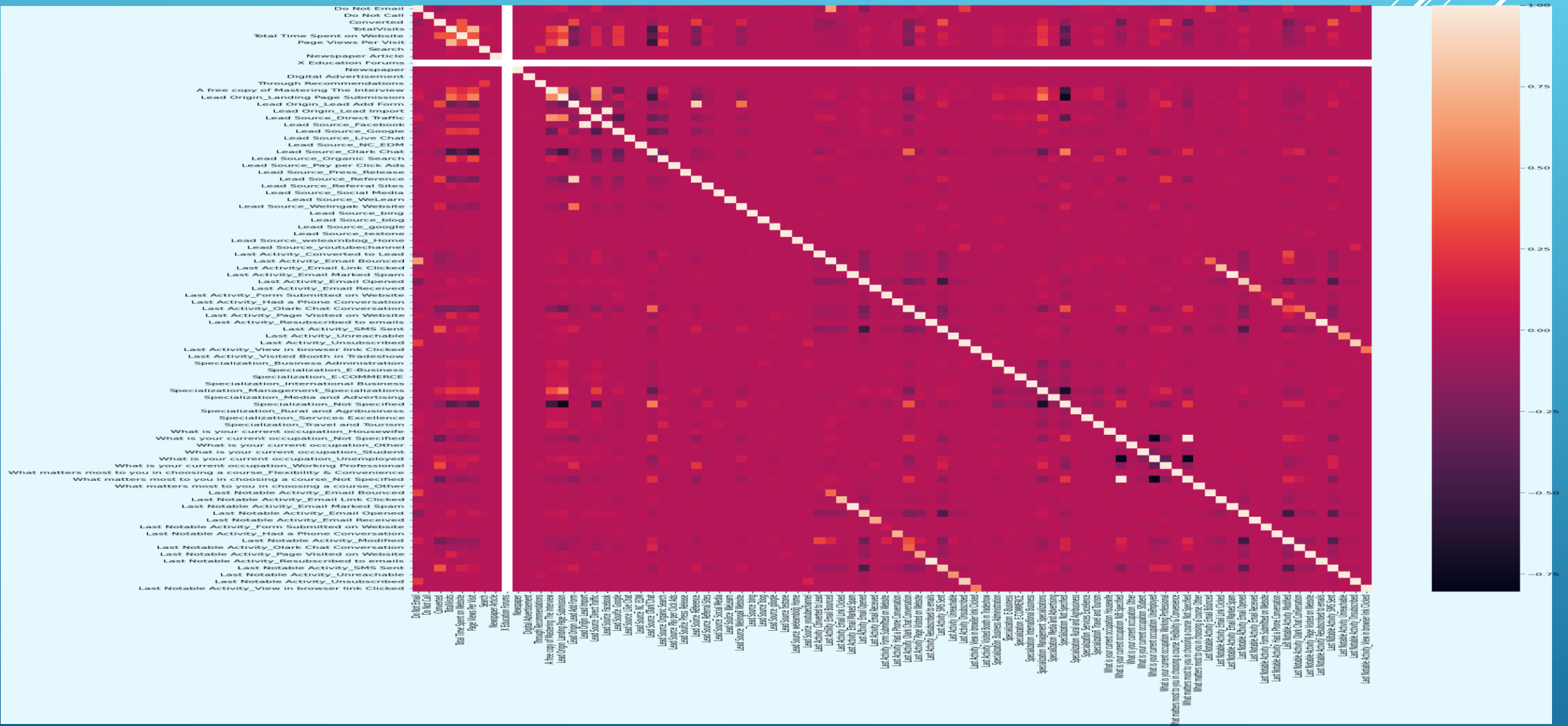
1.Total Visits

2.Total time spent on website

3.Page Views per visit
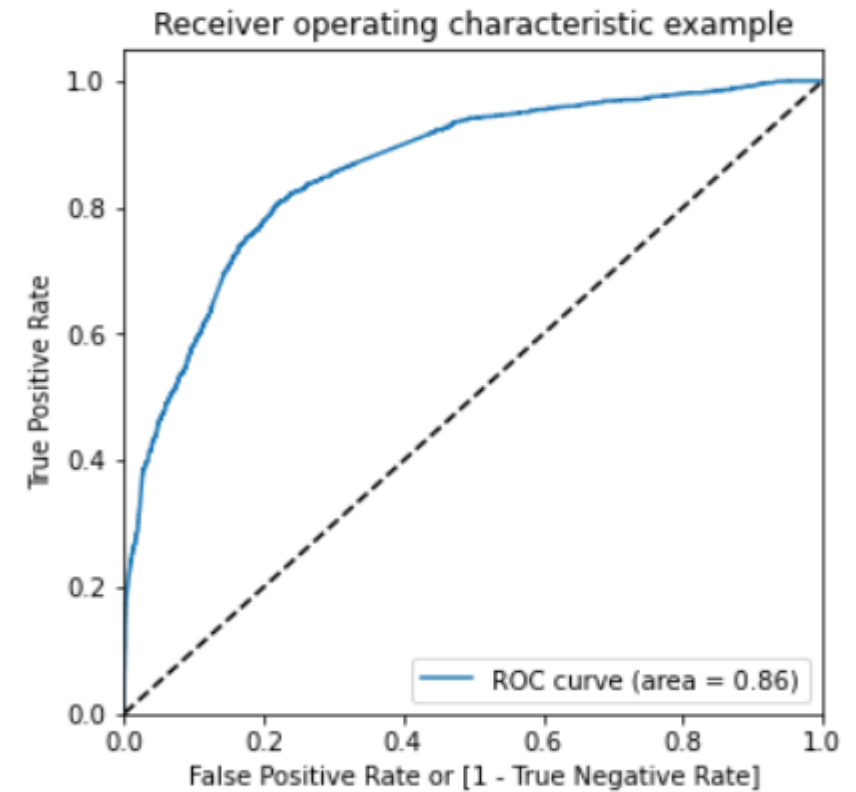
Correlation

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
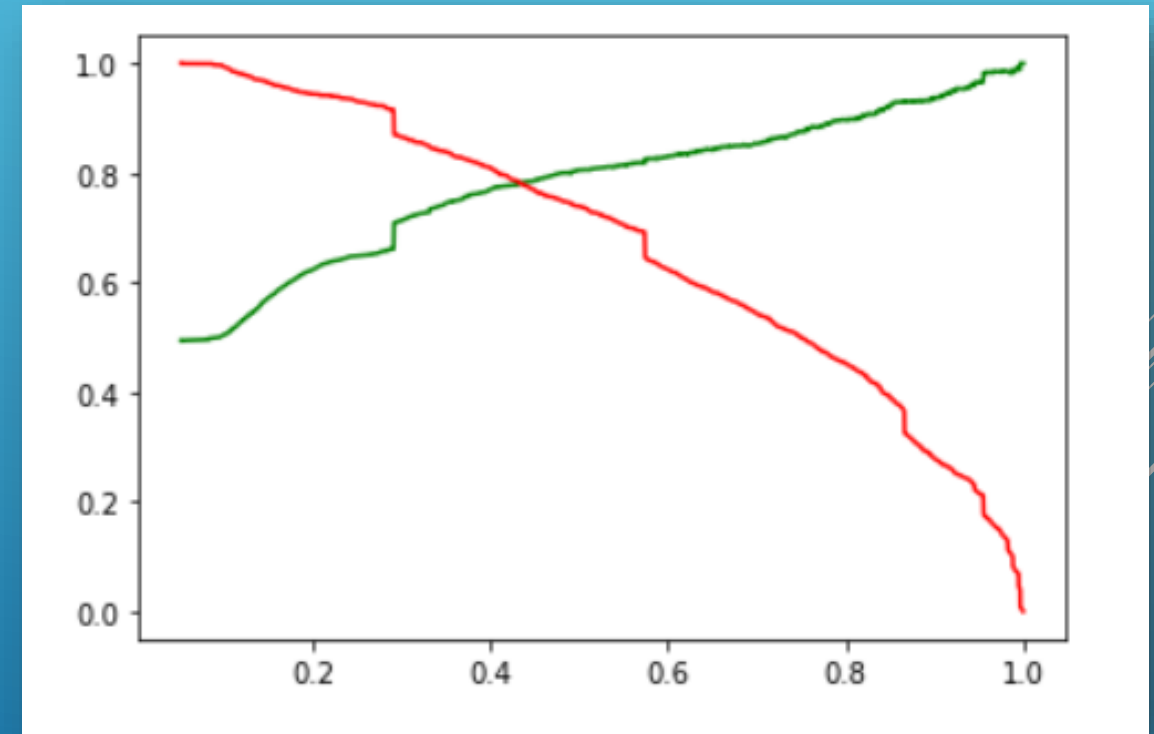
Accuracy, Specificity And Sensitivity

As you can see that around 0.2, you get the optimal values of the three metrics. So let's choose 0.2 as our cutoff now.
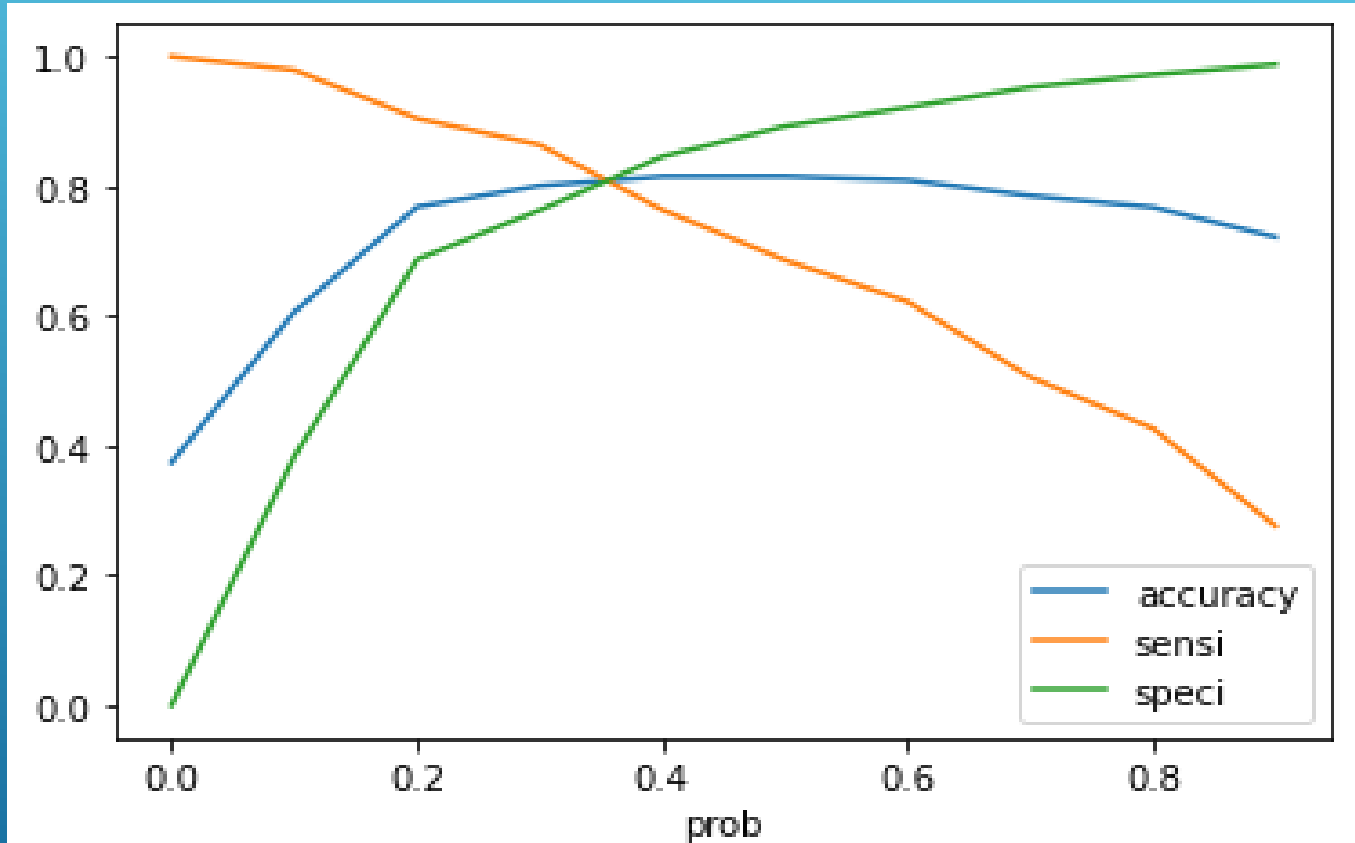
Precision Recall Curve

PREDICTION	TRAIN	TEST
PRECISION	0.78	0.78
RECALL	0.77	0.77

Conclusion

➢ There are many leads generated in the beginning (top), but only a small number of them turn into paying clients in the latter stages. To increase lead conversion, we have properly nurture the potential leads during the middle stage (e.g., by educating the leads about the product and maintaining ongoing communication).

➢ We sorted out the top prospects first from generated leads. The factors that have the biggest impact on the likelihood that a lead will be converted are "Total Visits," "Total Time Spent on Website," and "Page Views Per Visit." Then, we keep a list of leads close to hand so we can let them know about new programs, services, job openings, and upcoming higher education.

➢ Monitor each lead carefully so that we can personalize the content which will be send to them.

➢ Carefully present career opportunities, information, or training programs that best suit the leads' interests. A good strategy for identifying each lead's demands can help to convert leads into prospects.

➢ Concentrate on leads that have been converted.

➢ Engage leads in question-and-answer sessions to gather the pertinent data you require about them. To find out if the leads want to enroll in online courses, make more enquiries and appointments with them.