



Lead Score Case Study

MEMBERS:

- *Mohamed Thahir*
- *V Maya Jayachandran*
- *Shalini Jain*

PROBLEM STATEMENT

X Education sells online courses to industry professionals.

The company markets its courses on several websites and search engines like Google.

Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos.

When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals.

Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not.

The typical lead conversion rate at X education is around 30%.



**Problem
Statement**



BUSINESS GOAL:

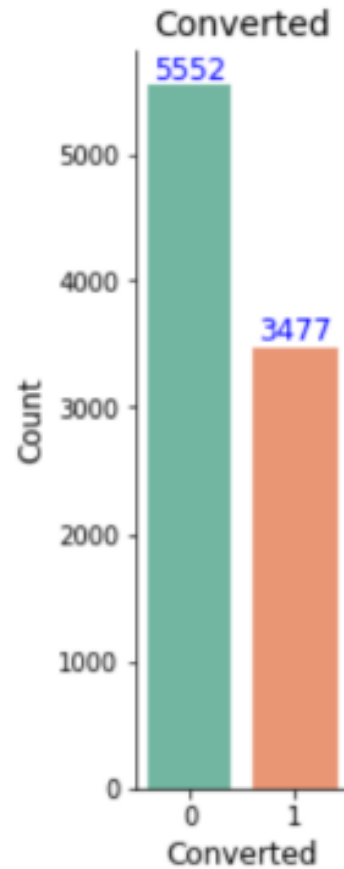
- X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

METHODOLOGY

- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

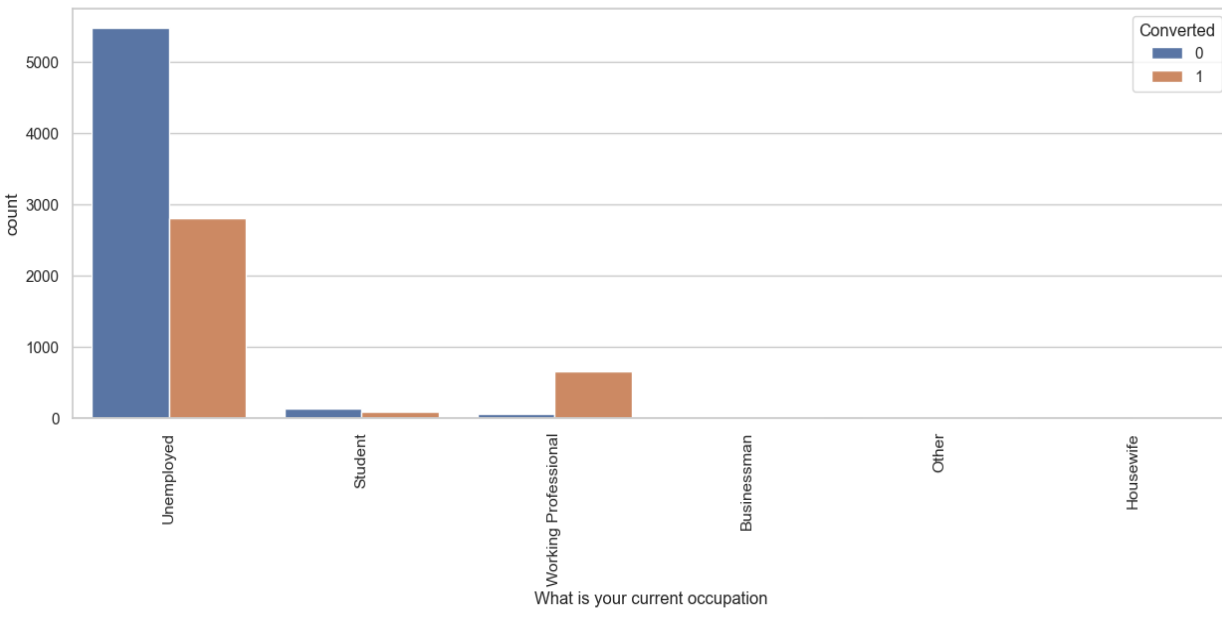
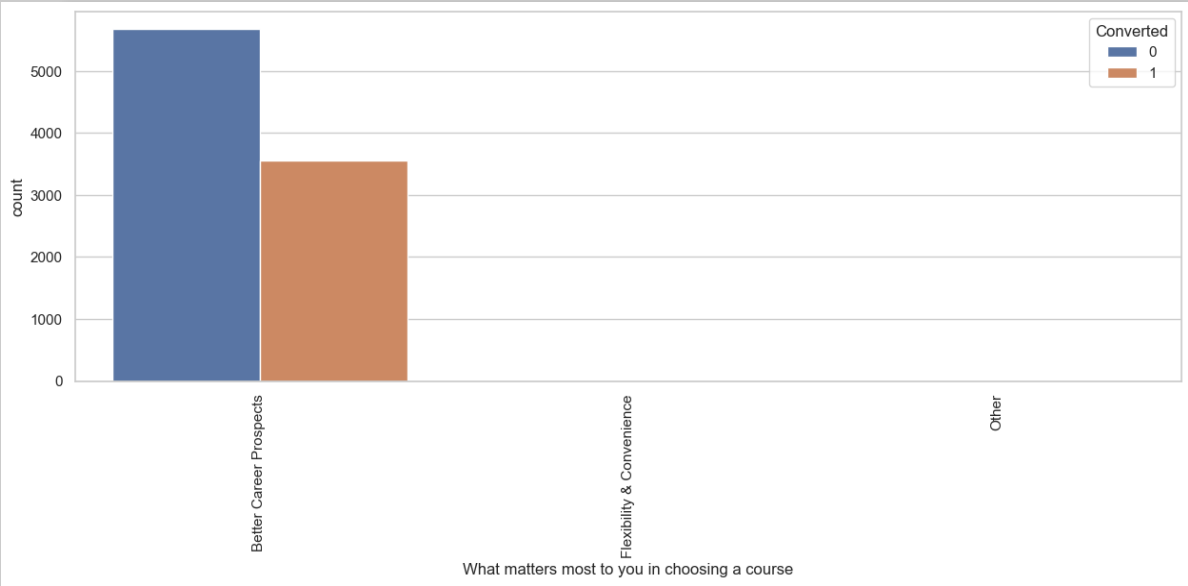
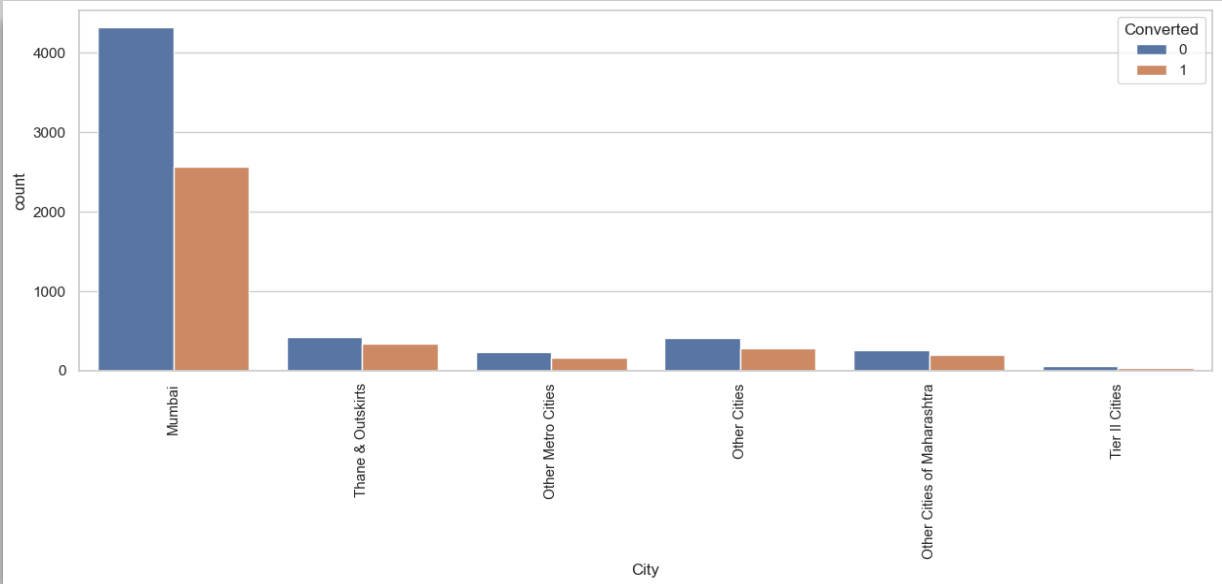
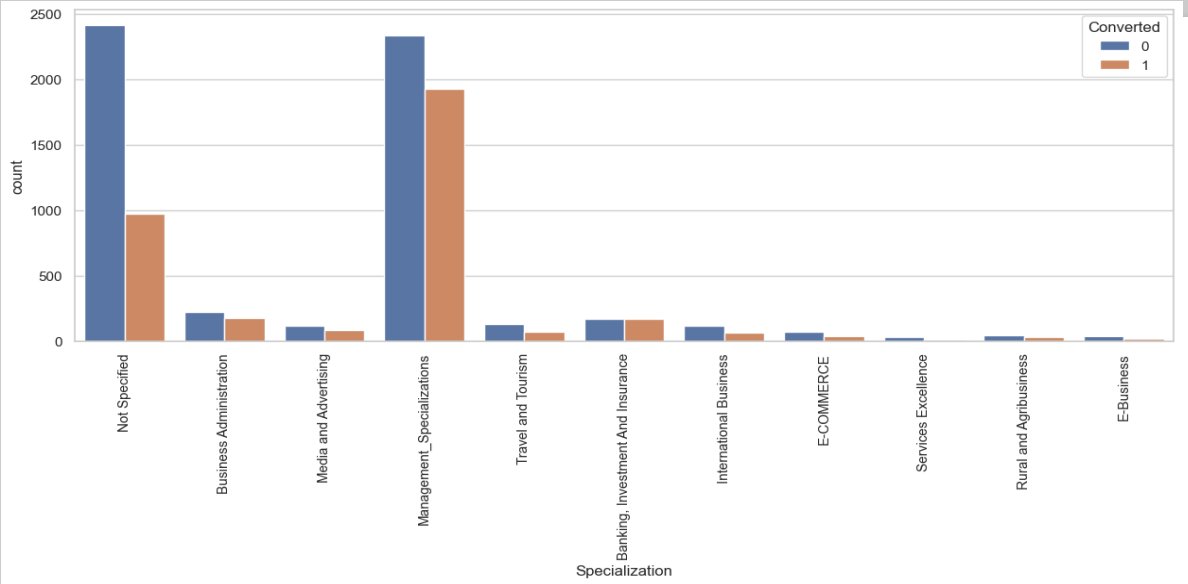


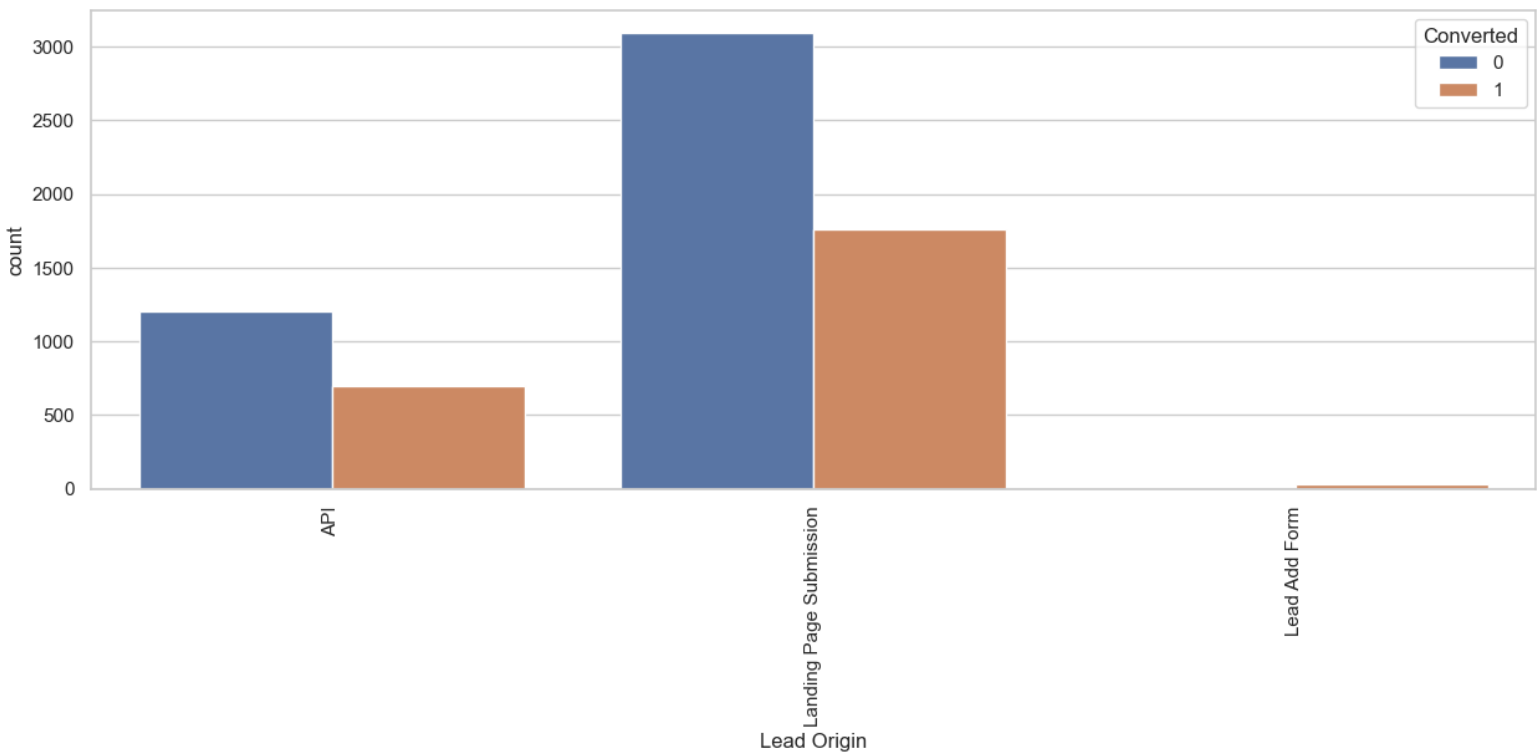
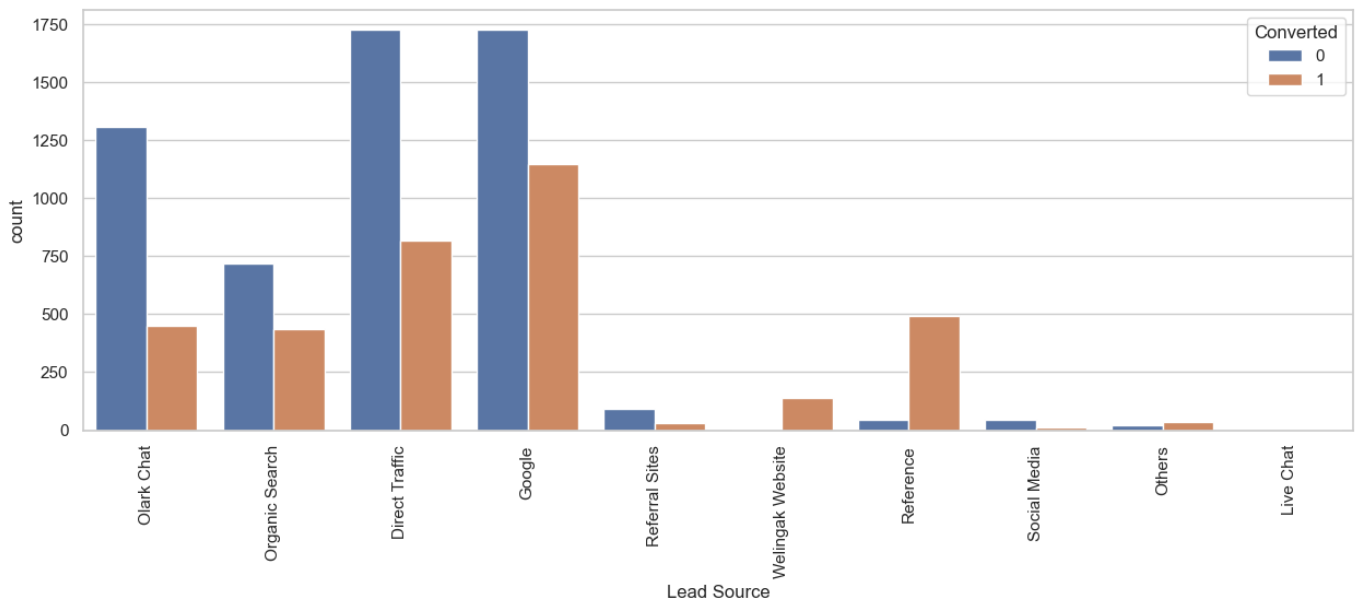
We have around 39% Conversion rate in Total



EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS





Following were the inferences obtained from the graphs through EDA

The maximum lead conversion was observed from the Lead Source **'Google'**.

The best conversion happens through the Lead Origin **'Landing Page Submission'**.

Most of the conversion have been made through the **Emails sent and the Calls being made.**

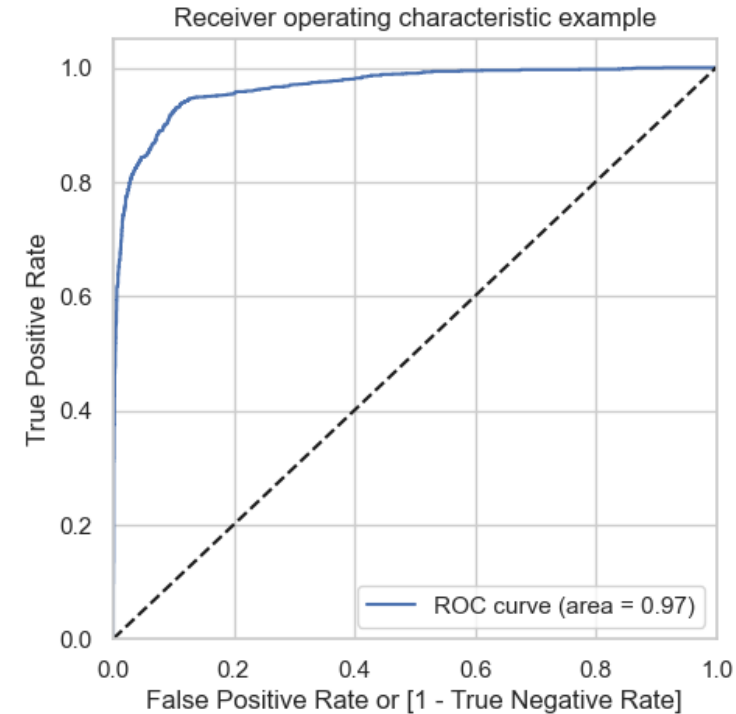
Most of the conversion is for people from the **'Management'** Specialization.

Also, majority of conversion are for people who live in the city **'Mumbai'**.

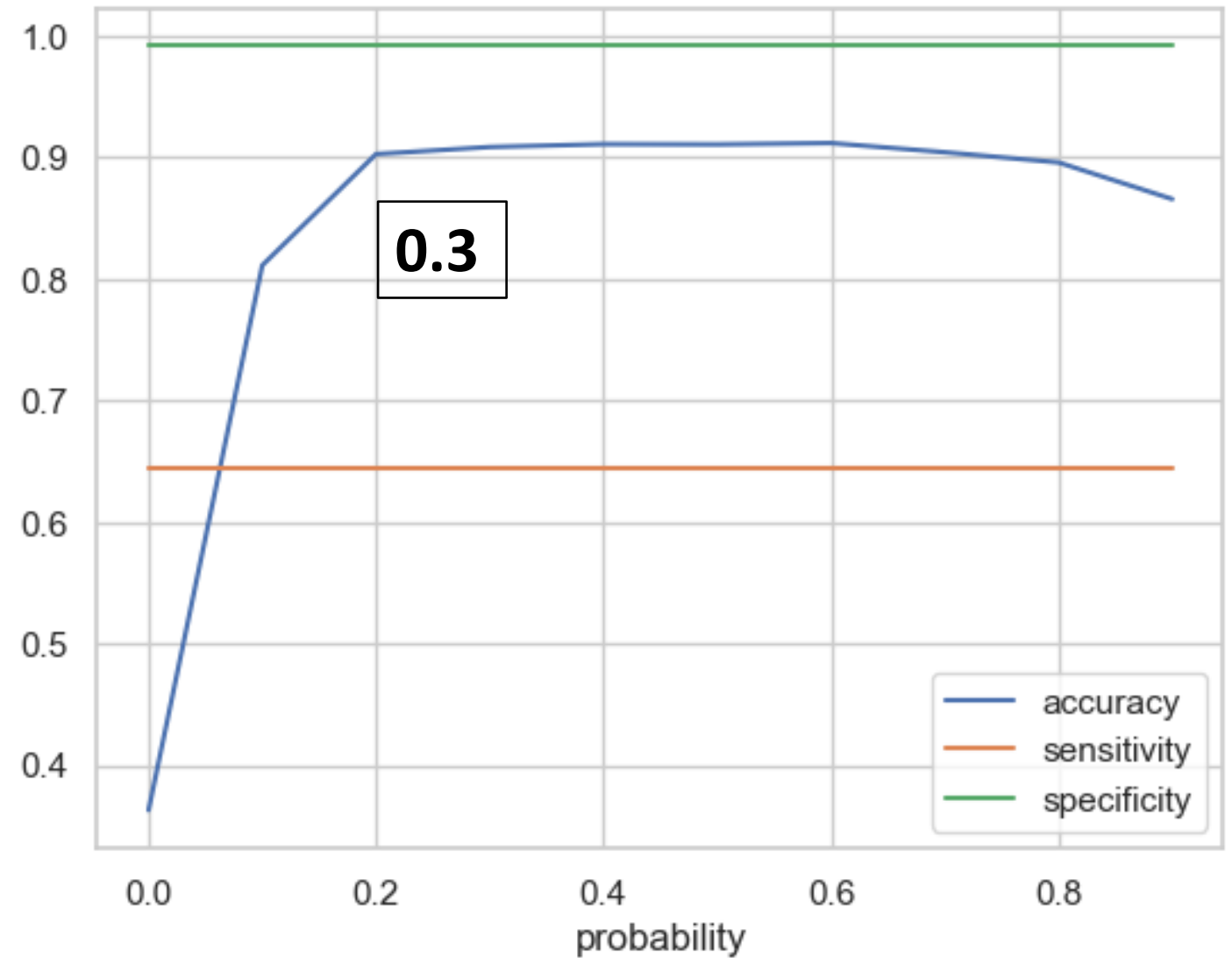
Most of the people are **'Unemployed'** who opt for conversions and their reason for opting for conversion is **'For better career prospects'.**

ROC Curve for the Train dataset

The Roc curve area should be close to 1, and our value is 0.97 which in turn indicates a good prediction model.



Cutoff
Probability
attained –
Training set



Measures attained : Training dataset



ACCURACY : 90.81%



SENSITIVITY : 92.05%



SPECIFICITY : 90.10%

Measures attained : Test dataset



ACCURACY : 90.62%



SENSITIVITY : 90.92%



SPECIFICITY : 91.41%

Observation:

The following figures are obtained after running the model on the Test Data:

Specificity: 90.62%

Accuracy: 90.92%

Sensitivity: 91.41%

Finally, consider the following: Let us compare the results for Train and Test:

Train Data:

Accuracy : 90.81%

Sensitivity : 92.05%

specificity : 90.10%

Test Data:

Accuracy : 90.92%

Sensitivity : 91.41%

Specificity : 90.62%

The Model appears to accurately anticipate the Conversion Rate, and we should be able to give the CEO confidence in making appropriate decisions based on this model.