

# Lead Score Case Study

**Submitted by :**

Prateek Raj Srivastava

Shreya Mandlesha

# Lead Score Case Study for X Education

## **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%.

## **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%.

# Strategy

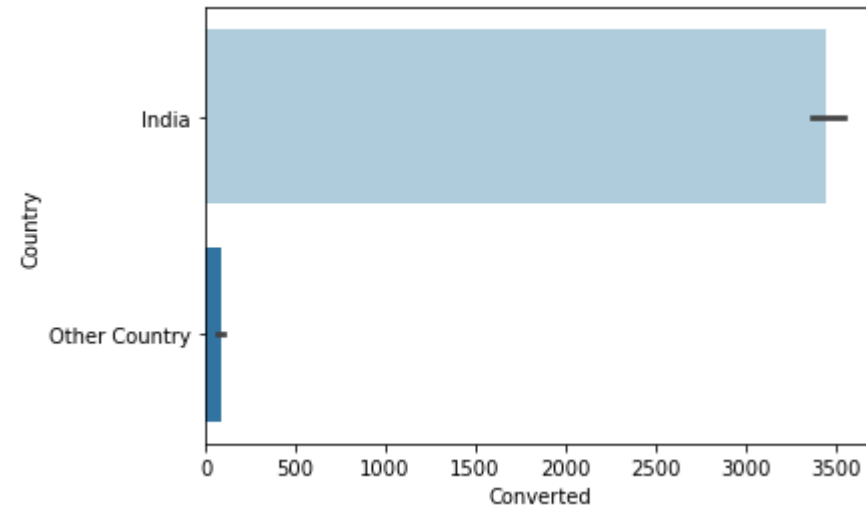
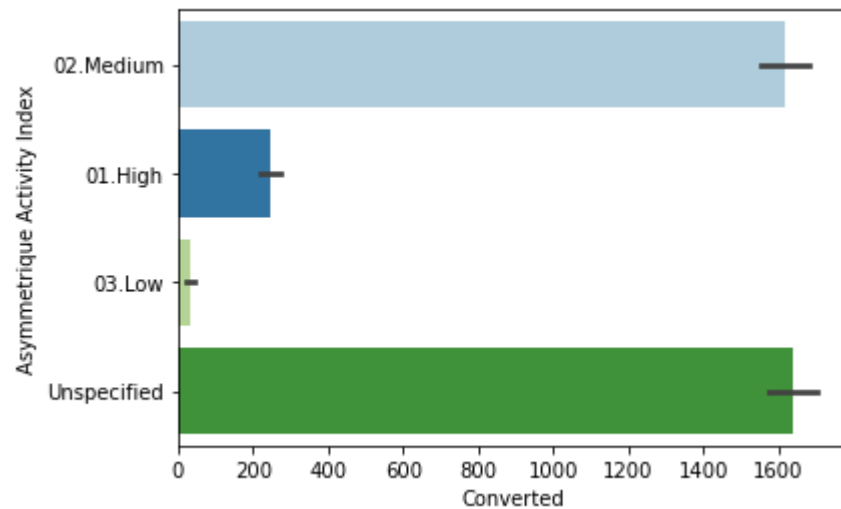
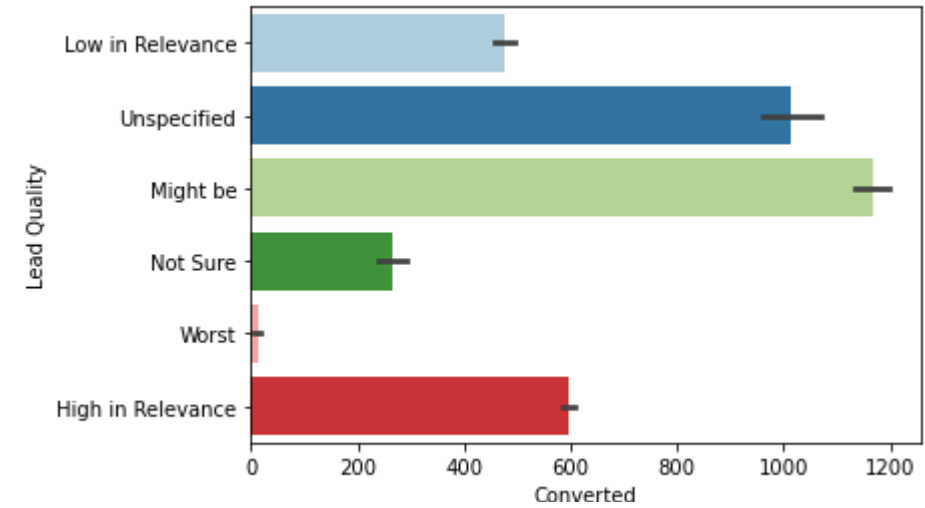
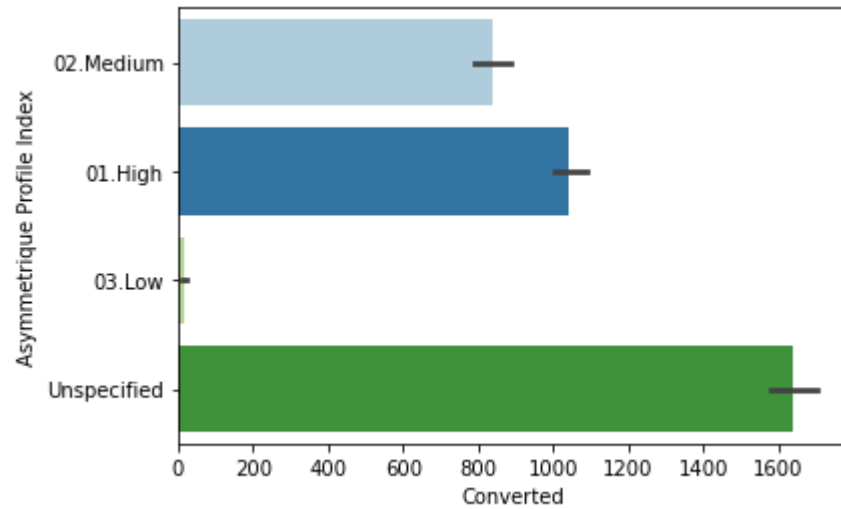
- ☐ Source the data for analysis
- ☐ Clean and prepare the data
- ☐ Exploratory Data Analysis.
- ☐ Feature Scaling
- ☐ Splitting the data into Test and Train dataset.
- ☐ Building a logistic Regression model and calculate Lead Score.
- ☐ Evaluating the model by using different metrics - Specificity and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

# Problem solving methodology

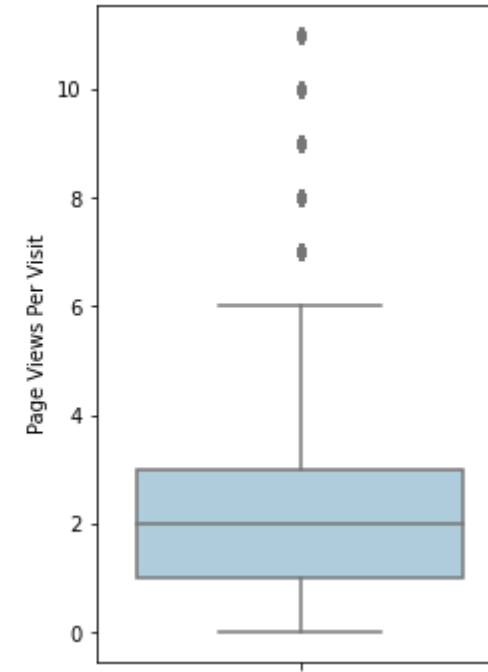
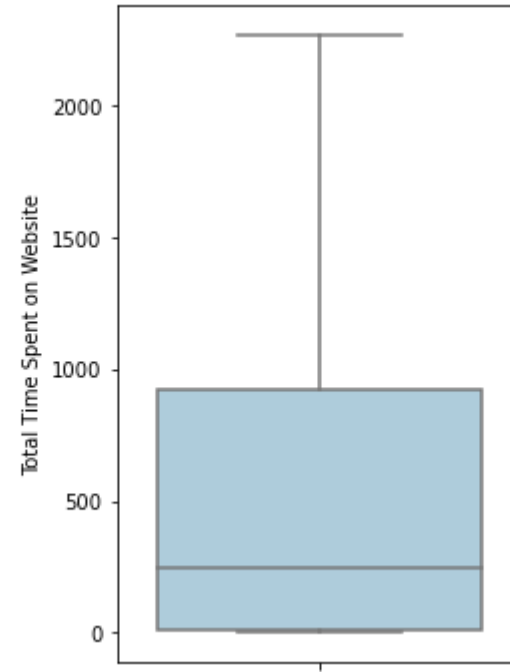
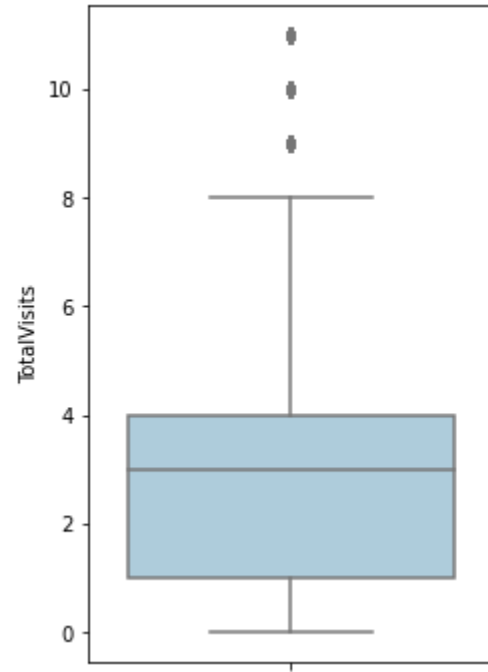


# Exploratory Data Analysis

We filled the null or missing values in a “unspecified” category



# Outlier Treatment

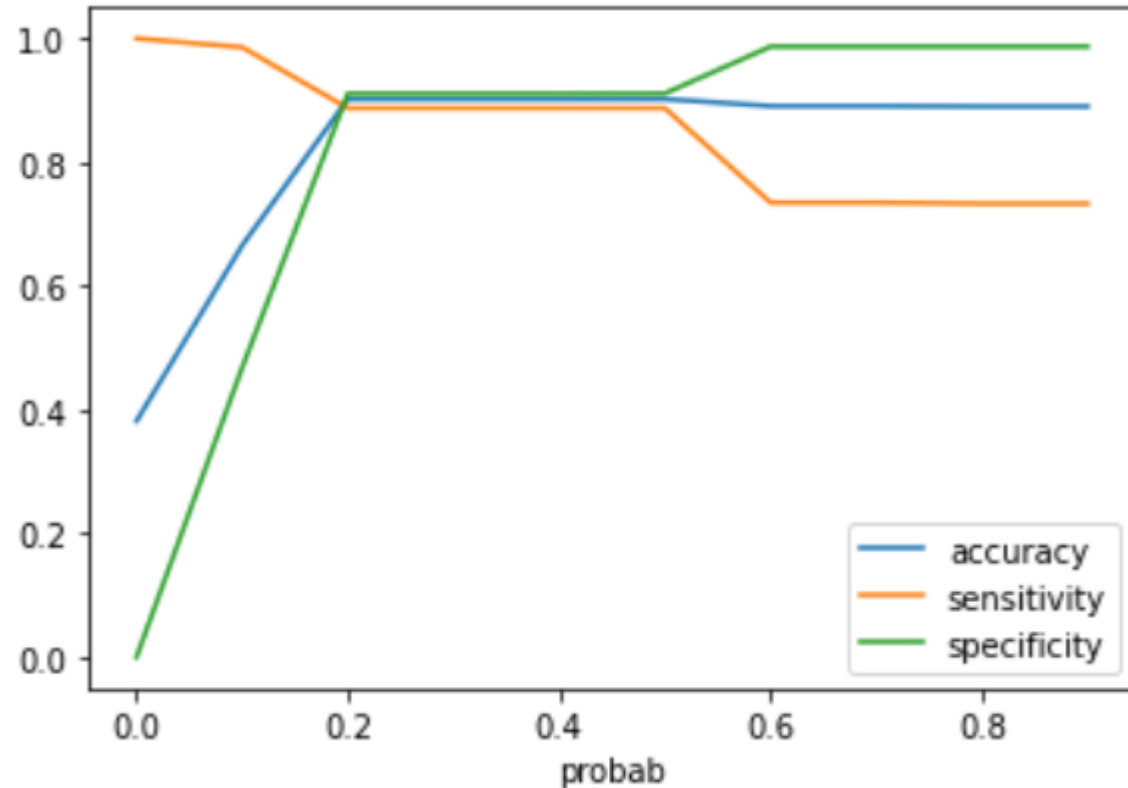


# Variables Impacting the Conversion Rate

- 'Lead Source\_Welingak Website',
- 'Lead Quality\_Worst',
- 'Tags\_Already a student',
- 'Tags\_Closed by Horizzon',
- 'Tags\_Interested in full time MBA',
- 'Tags\_Interested in other courses',
- 'Tags\_Lost to EINS',
- 'Tags\_Not doing further education',
- 'Tags\_Ringing',
- 'Tags\_Will revert after reading the email',
- 'Tags\_switched off',
- 'Last Activity\_SMS Sent'

## Model Evaluation - Sensitivity and Specificity on Train Data Set

we have took the cutoff 0.52

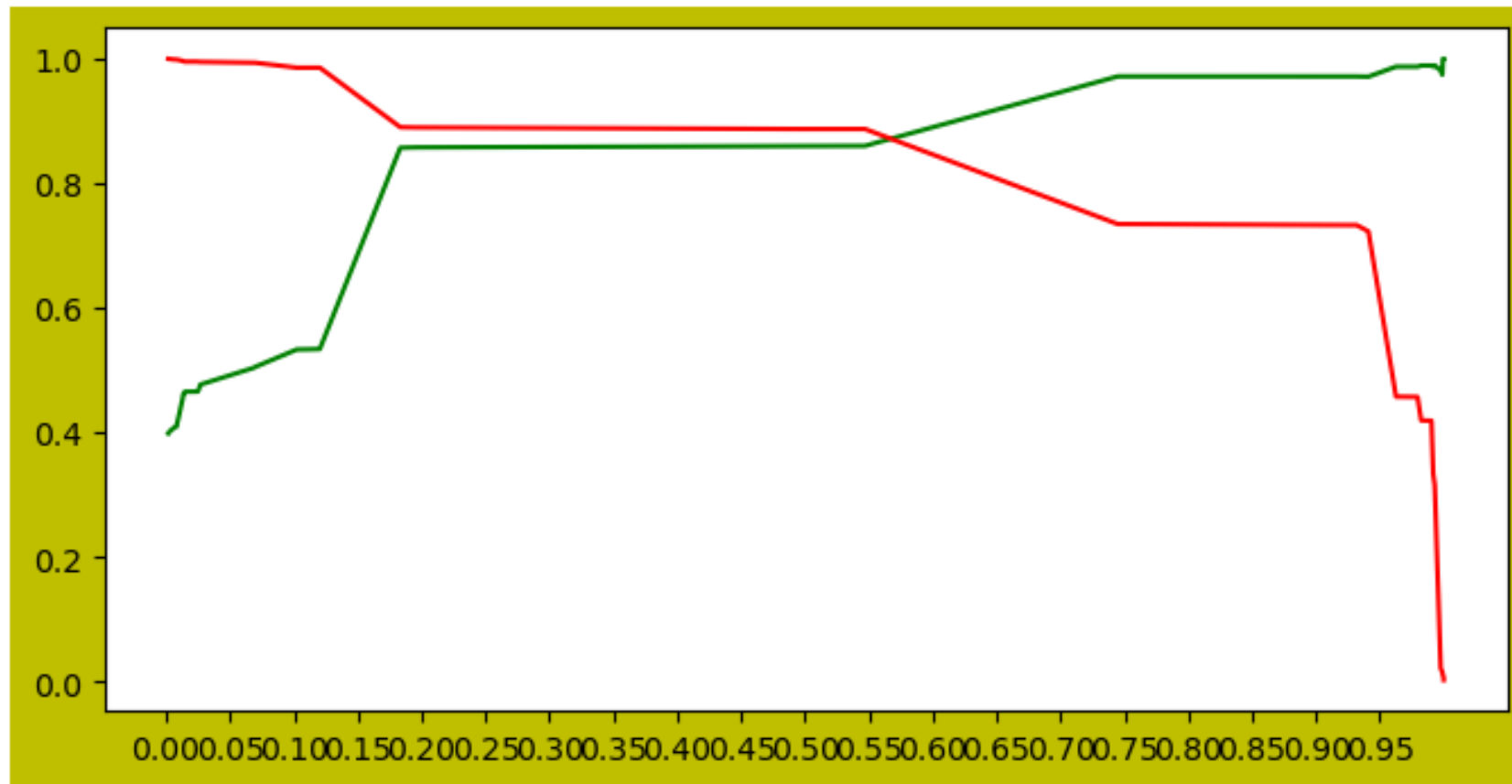


```
array([[3525, 344],  
       [ 269, 2117]], dtype=int64)
```

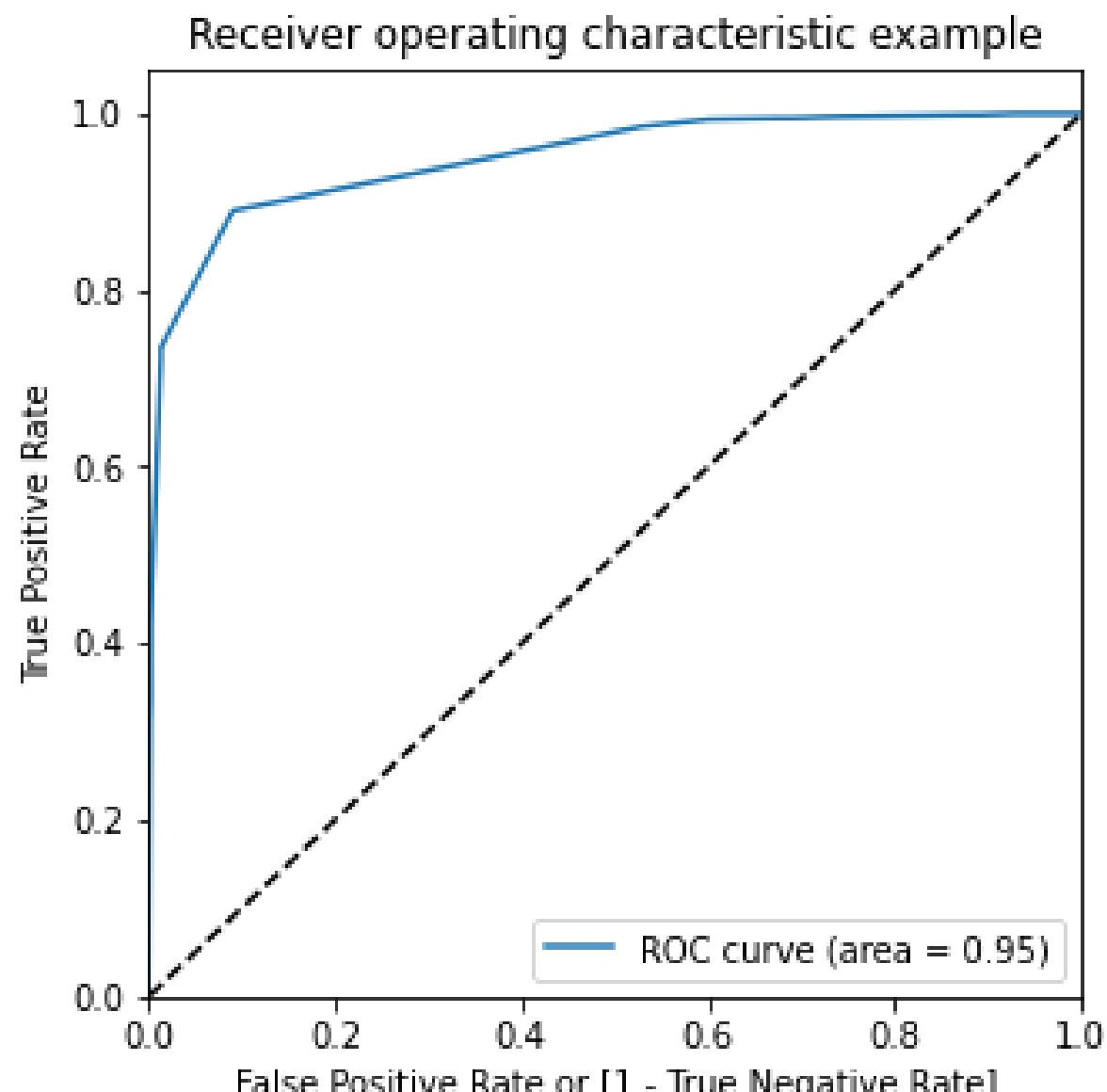


## Model Evaluation- Precision and Recall on Train Dataset

According to this the cutoff is near 0.57



# ROC Curve for train-data set



## Model Evaluation - Sensitivity and Specificity on Test Dataset

---

```
[[1500  153]
 [ 111  918]]
```

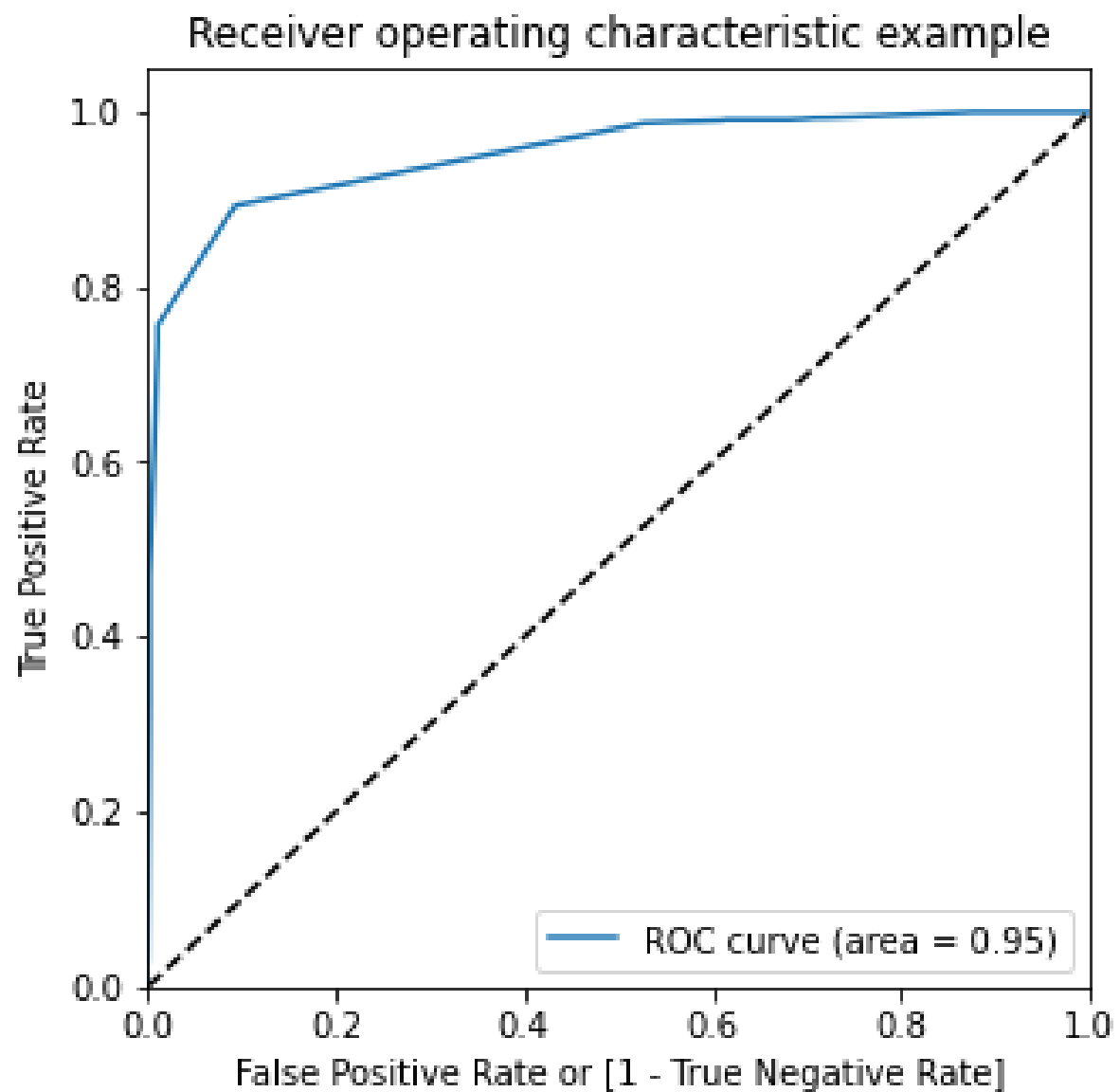
```
#Sensitivity
TP/float(TP+FN)
```

```
0.892128279883382
```

```
#Specificity
TN/float(TN+FP)
```

```
0.9074410163339383
```

## ROC Curve for test-data set



# Conclusion

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction. -
- Accuracy, Sensitivity and Specificity values of test set are around 90%, 89% and 90% which are approximately closer to the respective values calculated using trained set.
- Also the lead score calculated shows the conversion rate on the final predicted model is around 87% (in train set) and 87% in test set
- The top 3 variables that contribute for lead getting converted in the model are
  - Tags\_Lost to EINS
  - Tags\_Closed by Horizzon
  - Tags\_Will revert after reading the email
- Hence overall this model seems to be good.