# Lead Scoring Project

# Problem Statement

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

- 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%.

# Requirement

- The company requires a model for selecting promising leads based on lead score i.e assigned to each of the leads

-  such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.

- The ballpark of the target lead conversion rate has to be around 80% or more

# Strategy

1. Import data

2. Cleaning and preping the data

3. Performing Exploratory Data Analysis to understand the helpful features

4. Scaling the features

5. Seprating the data for train test split

6. Building a logistic regression model and assigning lead score

7. Testing the model on train dataset

8. Evaluating model using important metrics

9. Testing the model on test dataset

10. Measuring the accuracy and predictability of the model for evaluation

Focus on how we can improve the conversion rate of API and Landing Page Submission in order to attain a high conversion rate.
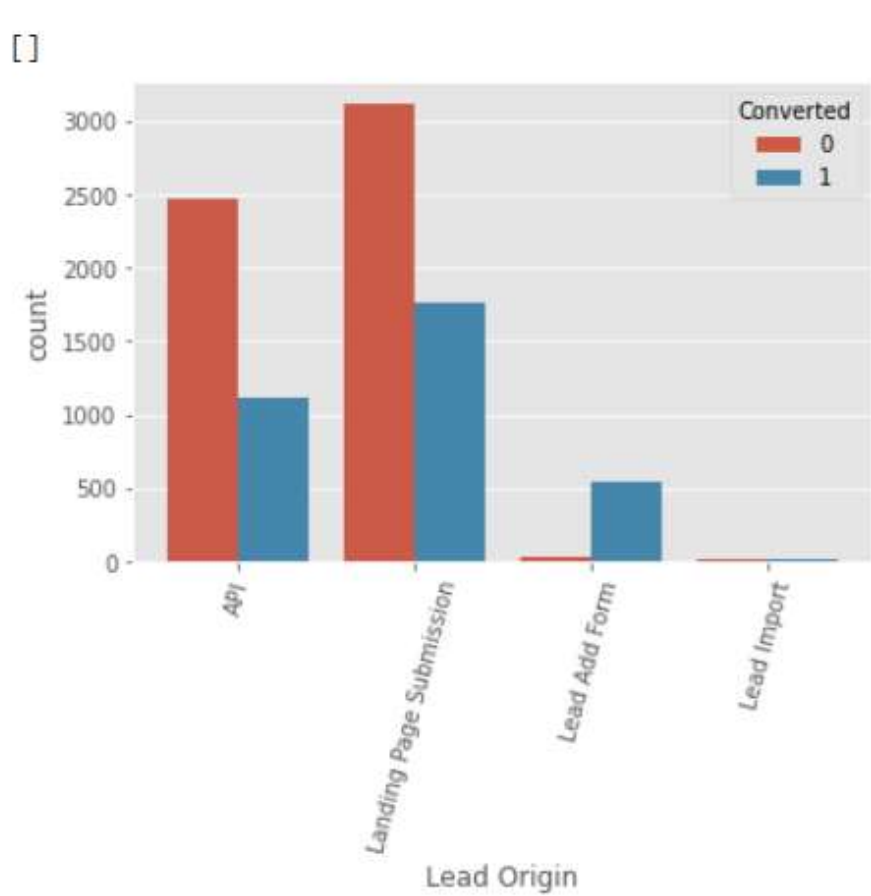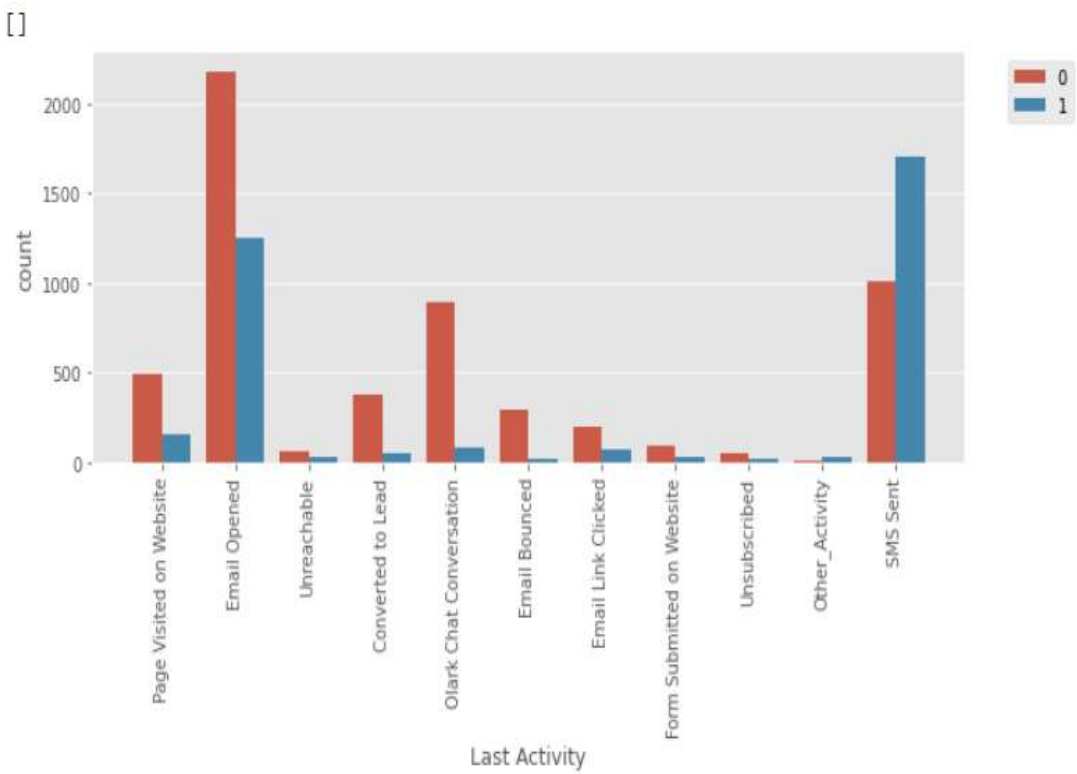
To improve the conversion rate We should Focus on improving the process of lead conversionfrom Olark Chat, Google Search, Direct Traffic, Organic Search. Also, we should try to generate more leads from reference and wellingak Website.
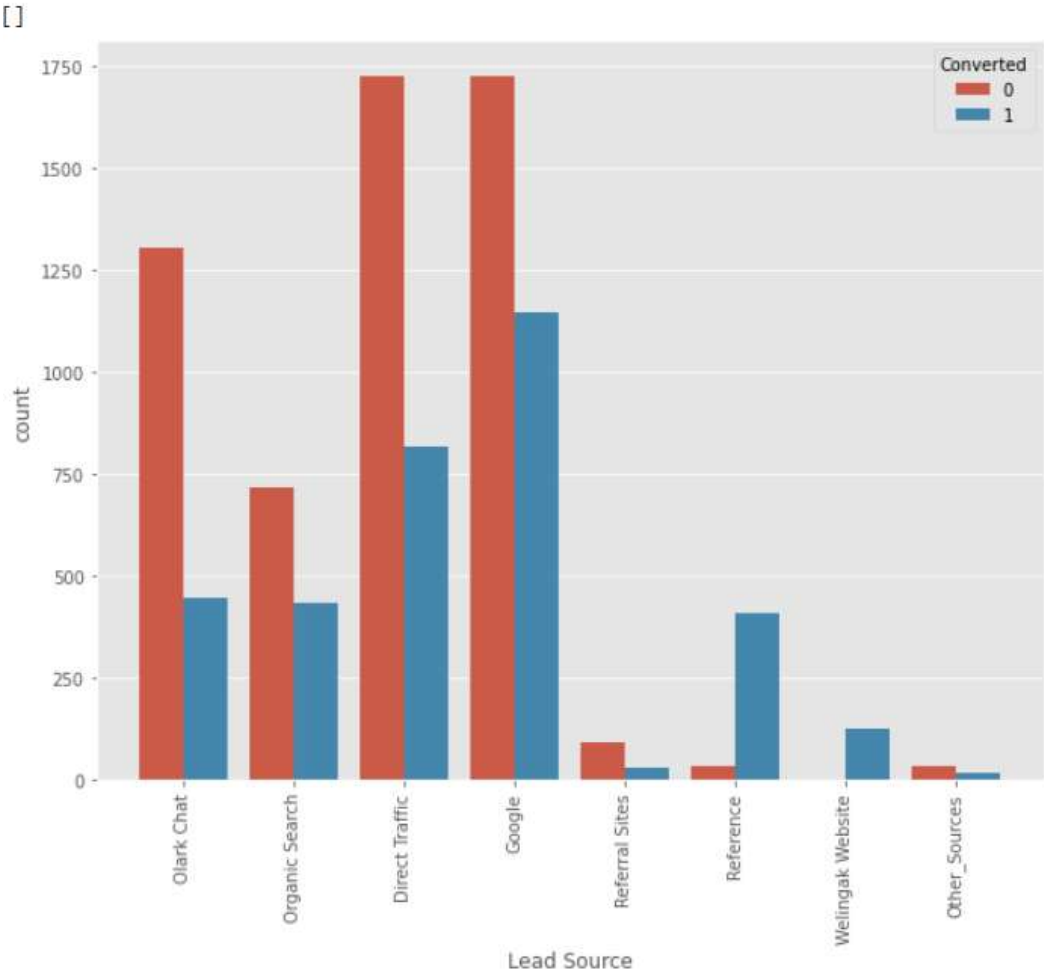
Total Time spent on website is an important factor to note and the leads converted are spending lot of time on the website. Hence website should be more engaging in order to get high conversion rates.
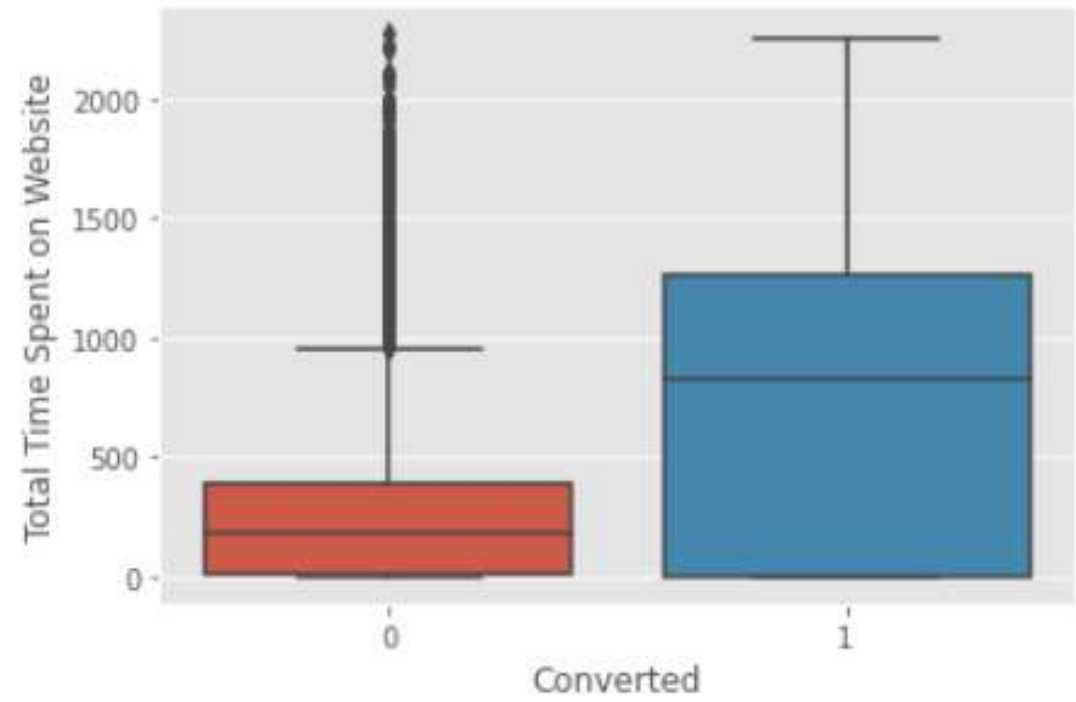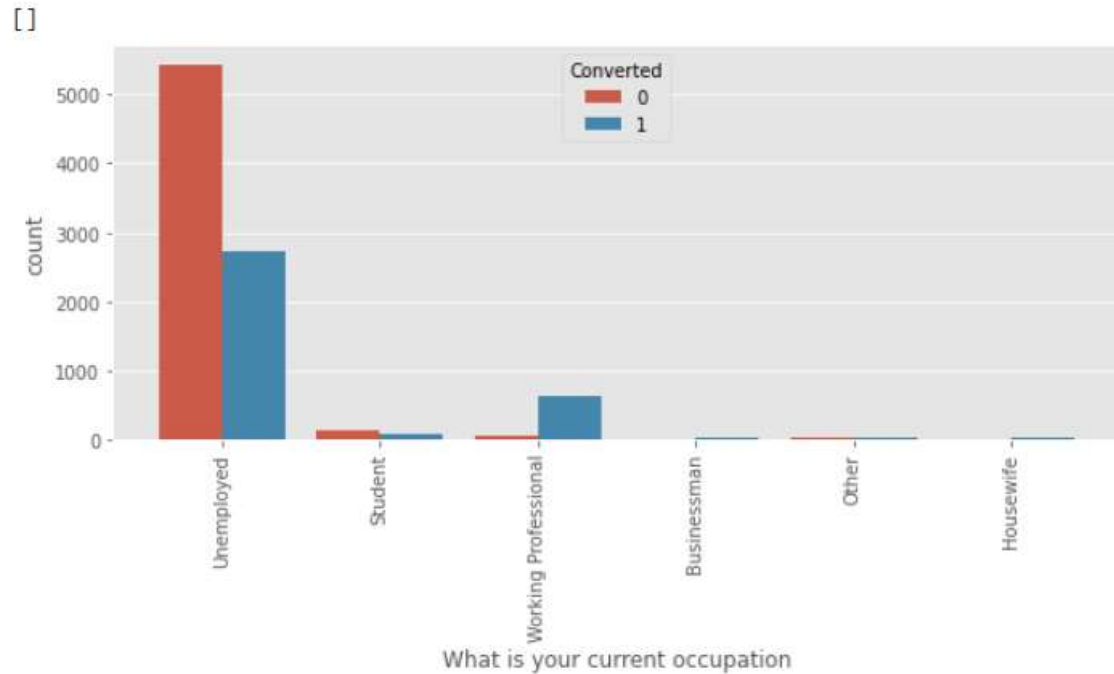
Conversion rate with last activity as SMS sent has a high conversion rate

Working Professionals have high conversion rate therefore the focus should be more on converting leads who are working.

Unemployed are high in number but the conversion rate is very low, to increase the overall conversion rate we should focus on converting unemployed leads

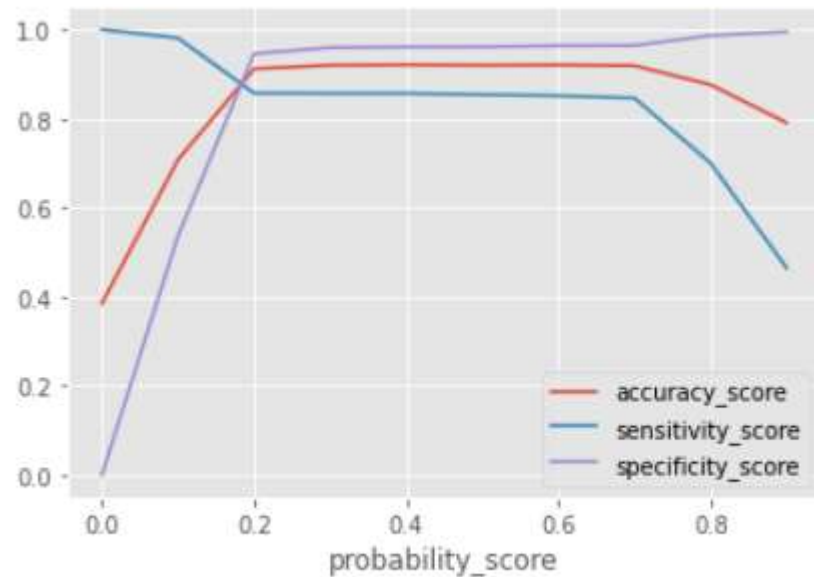Lead quality is an important metric. As it tells the type of lead.
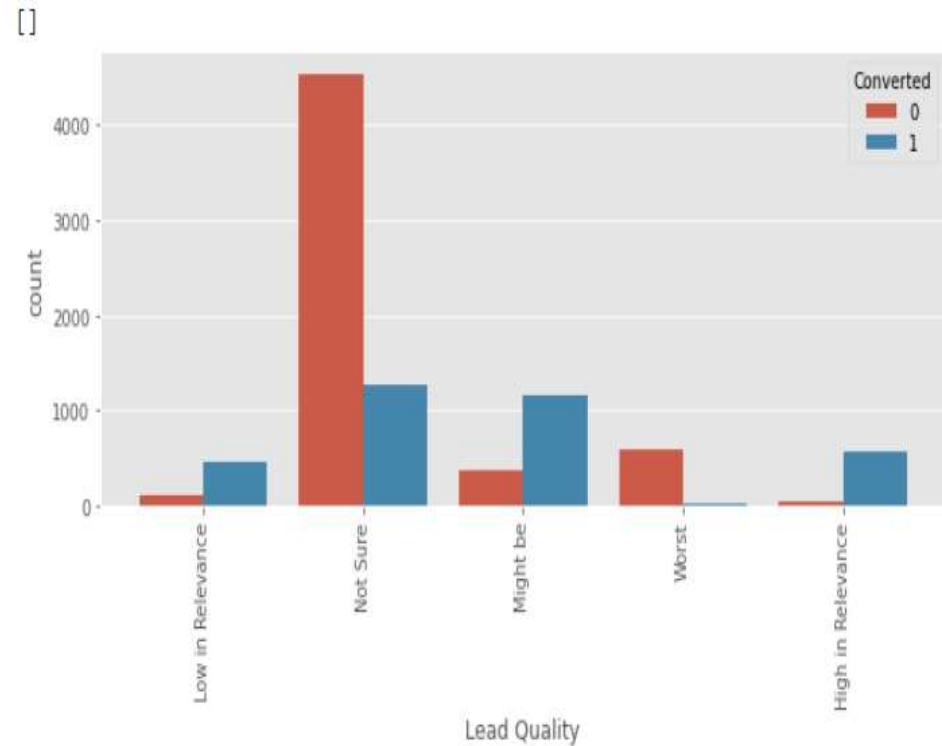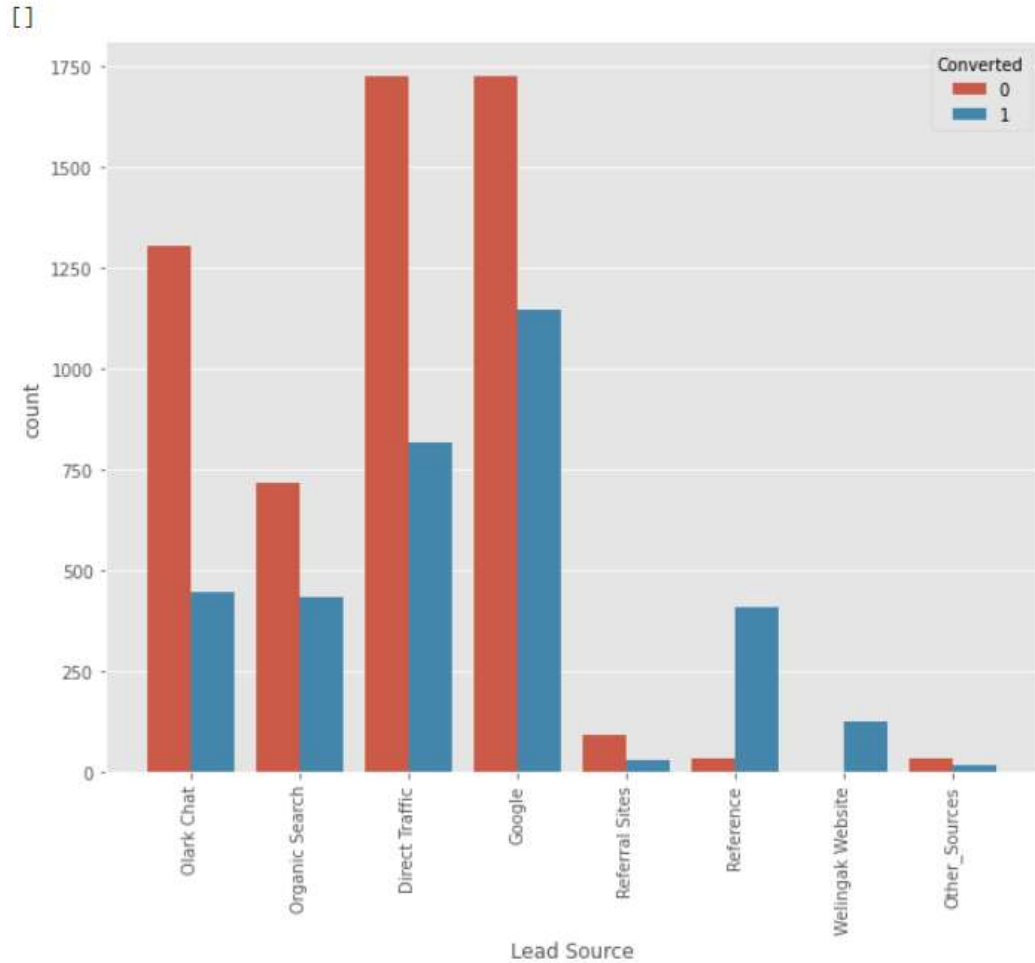
**Lead Source**
- Most of the Leads are coming from Olark Chat, Directly and with Google Search, Organic Search.

- Conversion rate of Reference and wellingak Website are high than other Sources.

- **Conclusion:**

To improve the conversion rate We should Focus on improving the process of lead conversionfrom Olark Chat, Google Search, Direct Traffic, Organic Search. Also, we should try to generate more leads from reference and Wellingak Website.

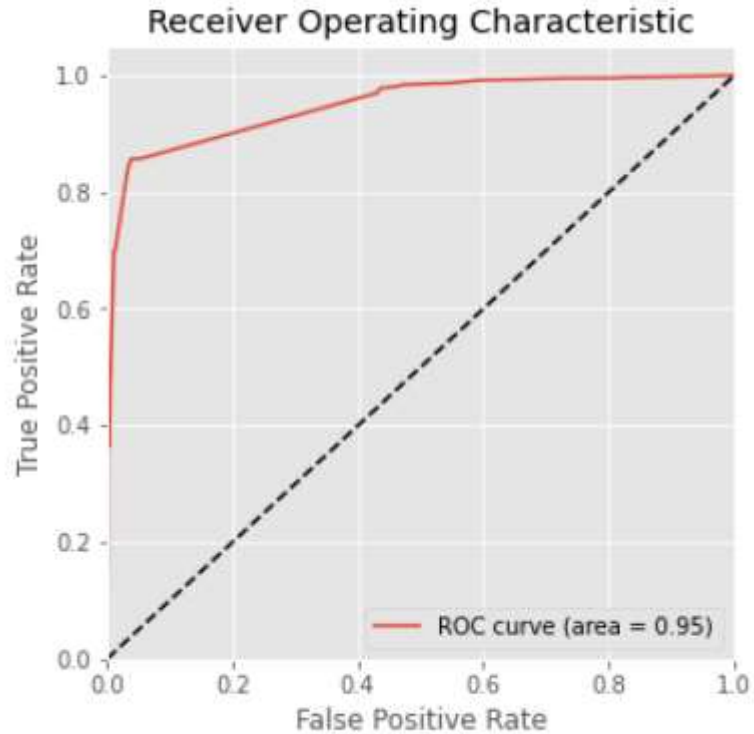## Training set Accuracy, Sensitivity and Specificity

```
                        Accuracy    : 0.91
[[3694  211]            Sensitivity : 0.85
 [ 349 2097]]           Specificity : 0.94
                        Precision   : 0.90
```

### Precision-Recall Trade off

# ROC CURVE



Receiver Operating Characteristic

AUC suggests no discrimination 0.7 to 0.8 is considered acceptable, **0.8 to 0.9** is considered excellent, and more than 0.9 is considered outstanding

# Test Data set Evaluation

```
Confusion Matrix : [[1642 92]
                    [ 157 832]]


Accuracy    : 0.90
Sensitivity : 0.84
Specificity : 0.94
Precision   : 0.90
```

## Conclusion

### EDA

1. People who are spending higher time on the website can be promising leads
2. SMS sent as the last activity have high impact on lead conversion
3. People looking for specialization in various fields have a high conversion rate.
4. Working Professionals have high chances of converting
5. Reference and offers for referring a lead can be a good source of higher conversion rate

### Logistic Regression Model:

1. Model shows high accuracy of more than 90%
2. Threshold have been selected from Precision recall curve and Accuracy, Sensitivity , Specificity Curve.
3. The model shows 84% sensitivity and 94% specificity
4. The model finds promising leads and is accurately predicts low and high converting leads