



# LEAD SCORE – CASE STUDY

By : Pratima  
Anjali Jha

# Problem Statement

- Company X Education sells online courses to industry professionals. The company gets the lead from various sources like Google, Past referrals etc.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc.
- The **Current conversion rate of the Company is 30%**
- **The company wants to identify the most potential leads, also known as 'Hot Leads' to increase the conversion rate.**
- sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.



# Analysis Steps:

## Data Cleaning

- Inspecting the dataset.
- 9240 records with 37 variables
- Null Values present in many features

## Null Value Treatment

- Converting Select to null value
- Dropping Variable with higher null value > 45%
- Imputing variable having low null values with mean / mode

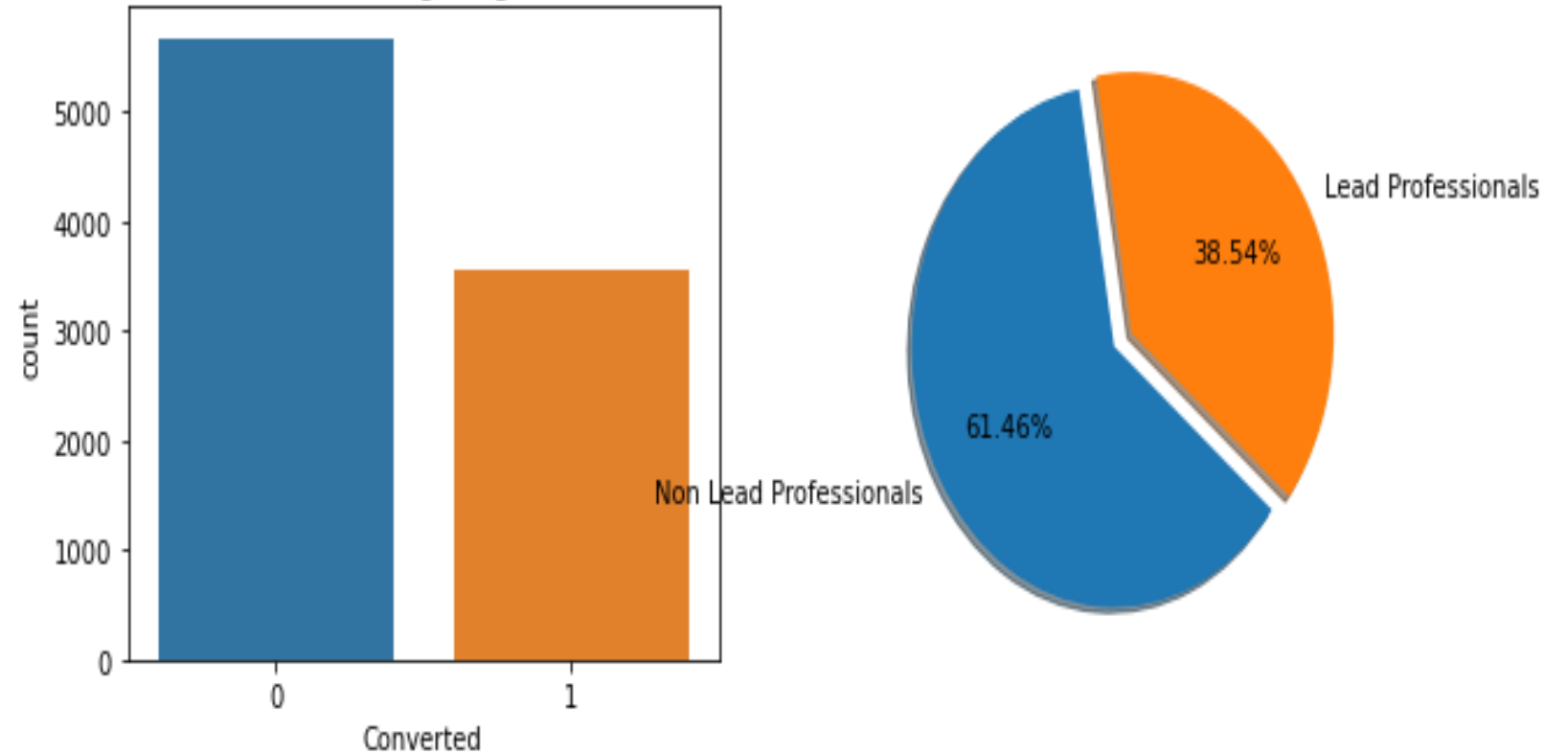
## Other cleaning

- Highly Skewed columns removed.
- Dropping Rows with more than 5 null values

# Performing EDA

## Distribution of Professionals getting enrolled in the course.

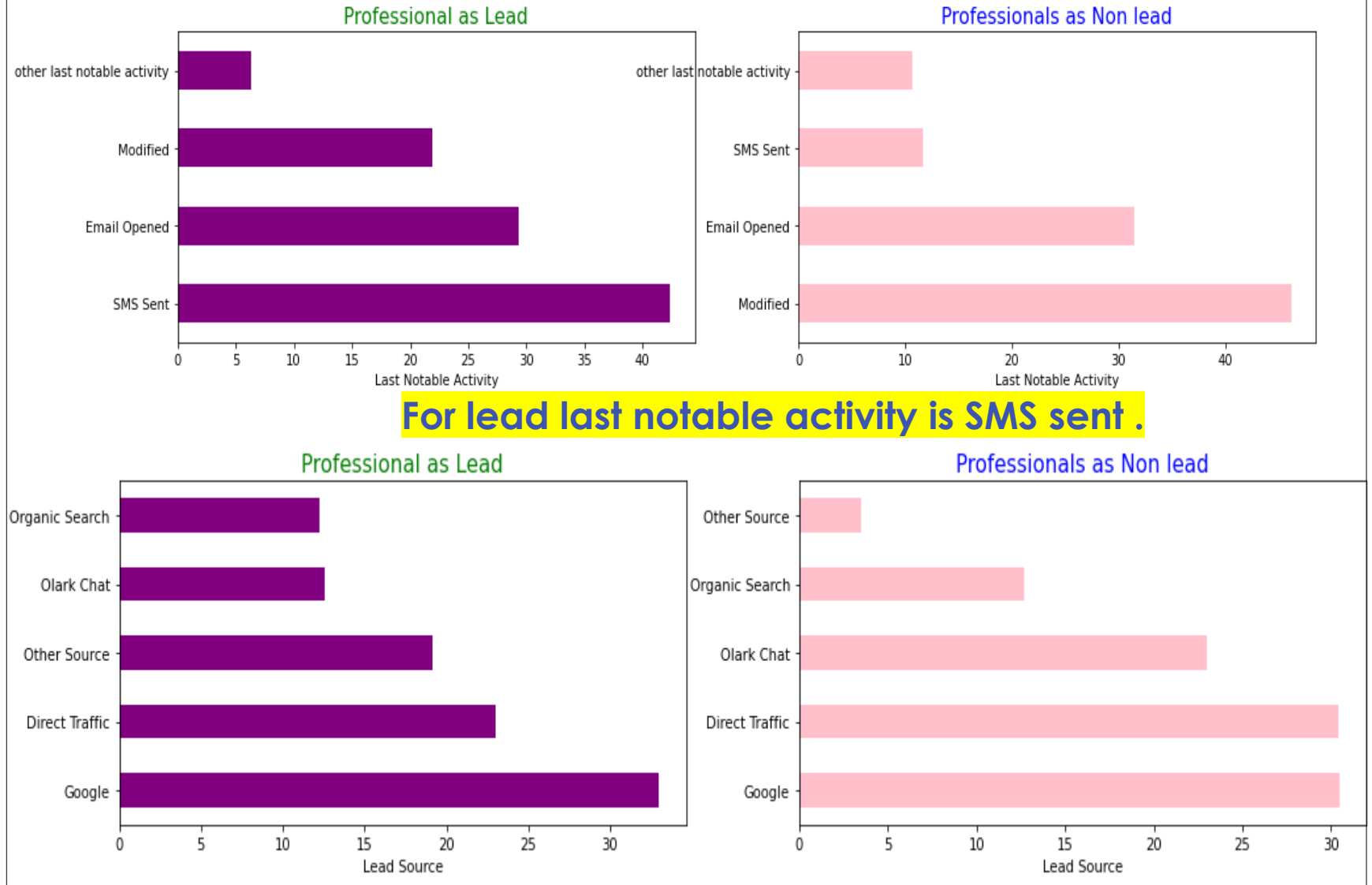
Distribution of Professionals of gettng enrolled in online coarse



About 38% of the leads are converted

# Performing EDA

## Univariate Analysis lead vs Non-Lead:

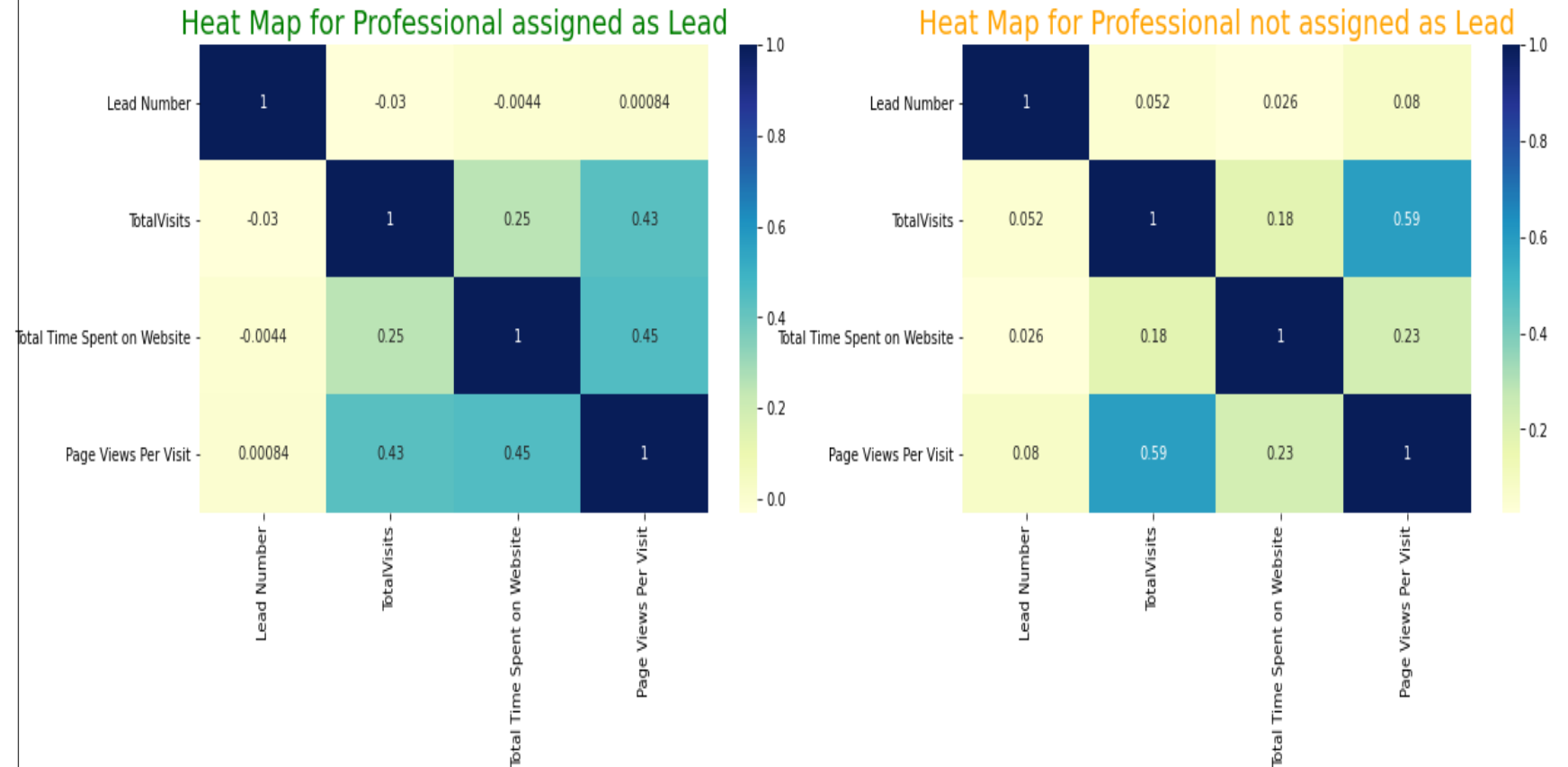


For lead last notable activity is SMS sent .

For Both Lead and Non Lead –Lead source is Google

# Performing EDA

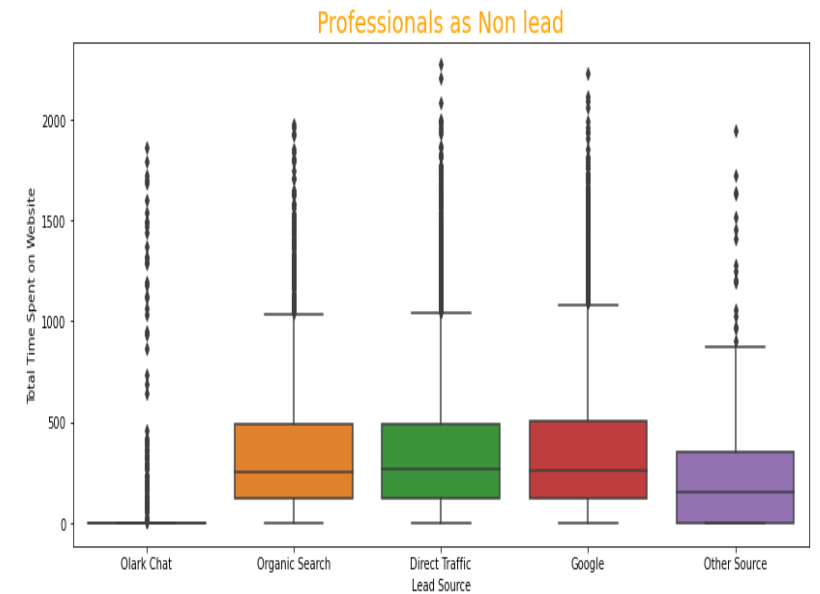
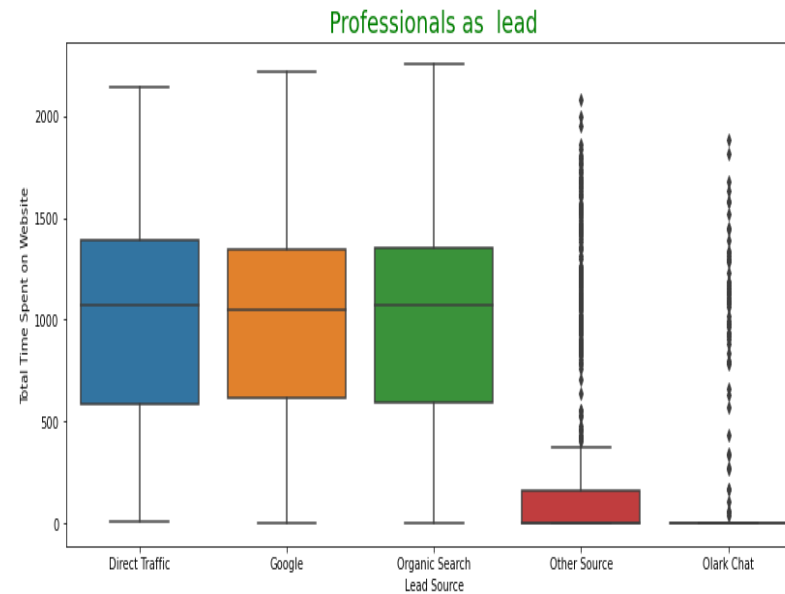
## Heat Map Lead Vs Non-Lead:



It can be observed that Total time Spent on Website is highly correlated with Total Visits and Page views per visits.

# Performing EDA

## Boxplot Lead Vs Non-Lead: Lead Source vs total time on website



It can be observed that Total time Spent on Website is more for converted lead in case the lead source is Direct Traffic or Google .

# Data Preparation & Modelling:

## Data Preparation

- Creating Dummy variables
- Data is converted to test train split
- Standardization of train data set using Standard-Scaler.

## Modelling

- RFE is used to choose 15 Top variables and other columns are dropped
- Using dual methodology P-Value and VIF to produce final model



# Modelling:

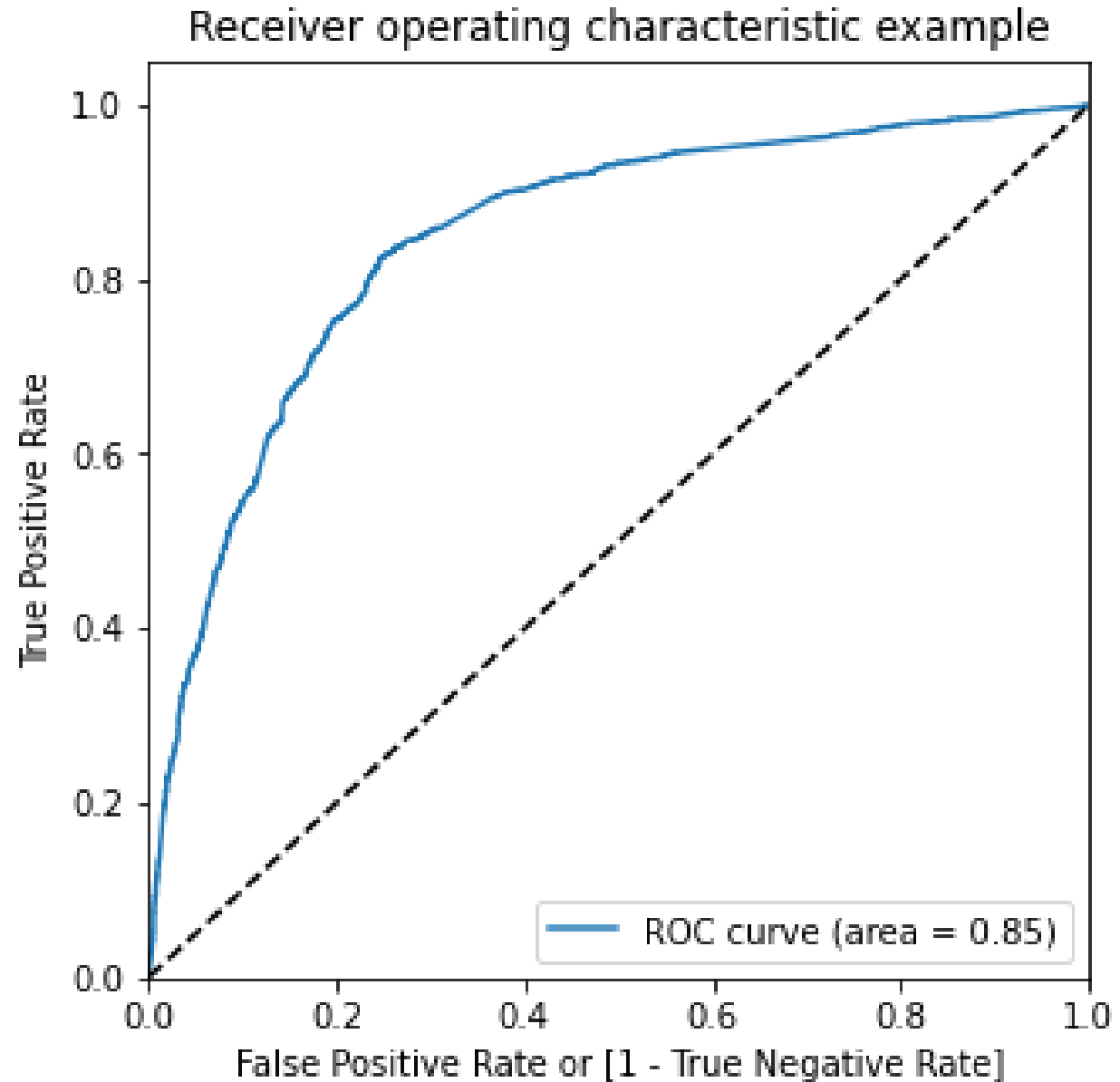
| Features         | Model 1                                | Model 2        | Model 3     | Model 4-<br>Final Model |
|------------------|--|----------------|-------------|-------------------------|
| Accuracy         | 78 %                                   | 77 %           | 77%         | 77.5%                   |
| P-Value          | High                                   | High           | High        | Low                     |
| VIF Value        | High                                   | High           | Low         | Low                     |
| Variable Dropped | LeadOrigin_Landin<br>g Page Submission | LeadOrigin_API | Total Visit | -                       |

# Sensitivity, Specificity for Model 4

- **Sensitivity**: It is True positive divided by total positives. It tells us how accurately the model has predicted 1 ie how accurately model has predicted the leads . Our target is to maximize the sensitivity.
- In Model 4 we got about 84% Sensitivity.
- **Specificity**: It is True negative divided by total negatives. It tells us how accurately the model has predicted 0 ie how accurately model has predicted the non-leads .
- In Model 4 we got about 73.5% Specificity.

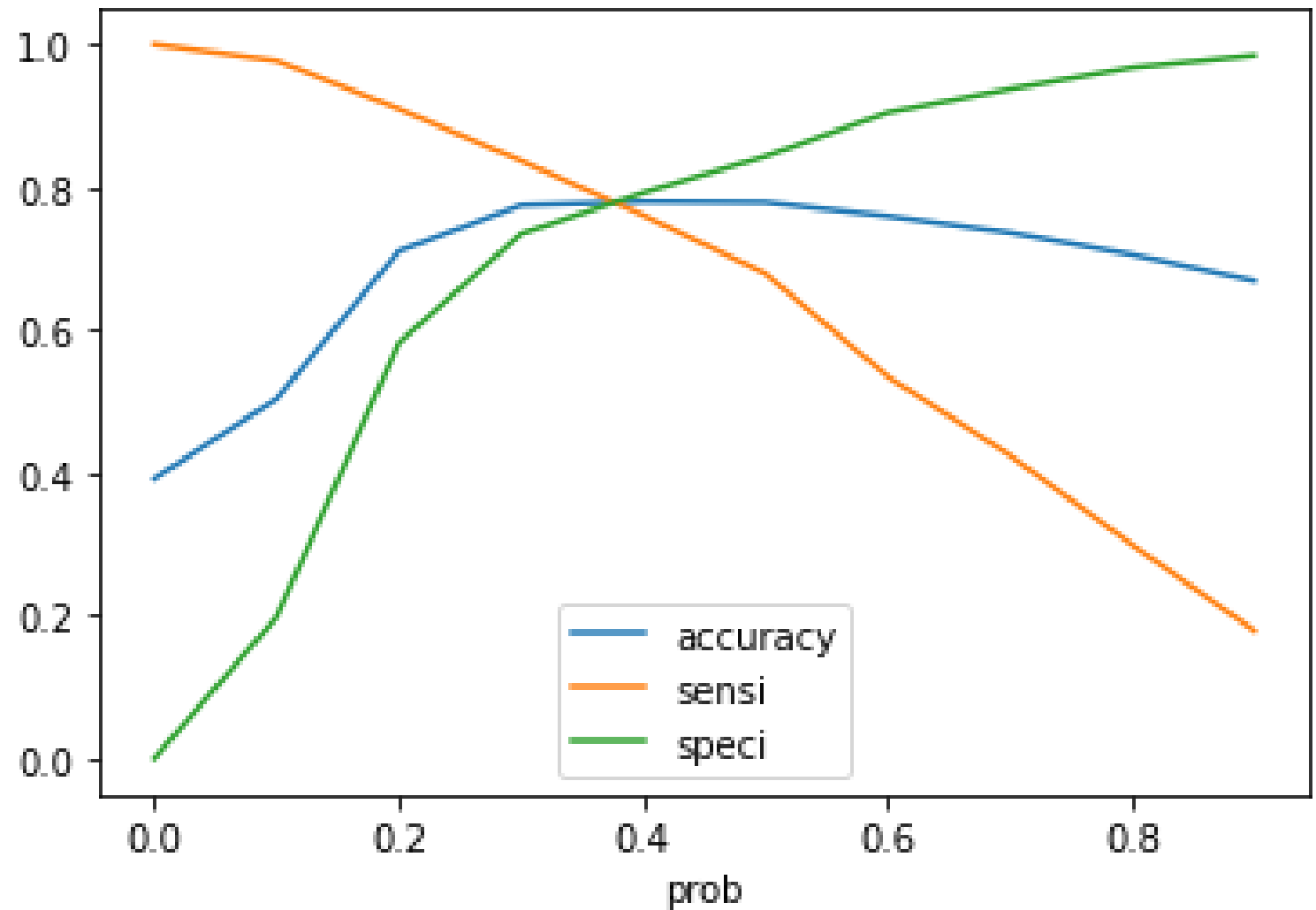
# Roc Curve

- It is the plot between false Positive rate and true positive rate . As the ROC Value increase the model performs better.
- The ROC of the model is 85%



## Understanding the Optimum Threshold:

- From the curve can be seen that , 0.3 is optimum cut-off.
- The Precision is 67% and Recall is 83% for the model.



# Making predictions on Test set

- The Accuracy of the model in test data is about 77% with sensitivity of about 84% and specificity of 72.4% .
- The Lead score is generated for the leads (range 0-100) using the probability of the lead to be converted.

A photograph of a business handshake, overlaid with a white double-lined diamond shape. The text 'Thank you' is centered within the diamond. The background is a blurred office setting with people in business attire. A document with the word 'CONTRACT' is visible in the lower right.

Thank  
you