

# LEAD SCORE CASE STUDY

Assignment Group Members : Zeel, Vijyalakshmi, Vyshnav

# Problem Statement

## Introduction:

- ❖ An X education company sells online courses to industry professionals. The company markets its courses on various websites and search engines such as Google.
- ❖ Once 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. The typical lead conversion rate at X education is around 30%.

# Business Goal

- ❖ Company wishes to identify the most potential leads, also known as “Hot Leads”.
- ❖ The company needs a model wherein a lead score is assigned to each of the leads such that the customer with higher lead score have a higher conversion chance and customer with lower lead score have a lower conversion chance.
- ❖ The CEO, in particular, has given a ballpark number for the lead conversion rate i.e. 80%

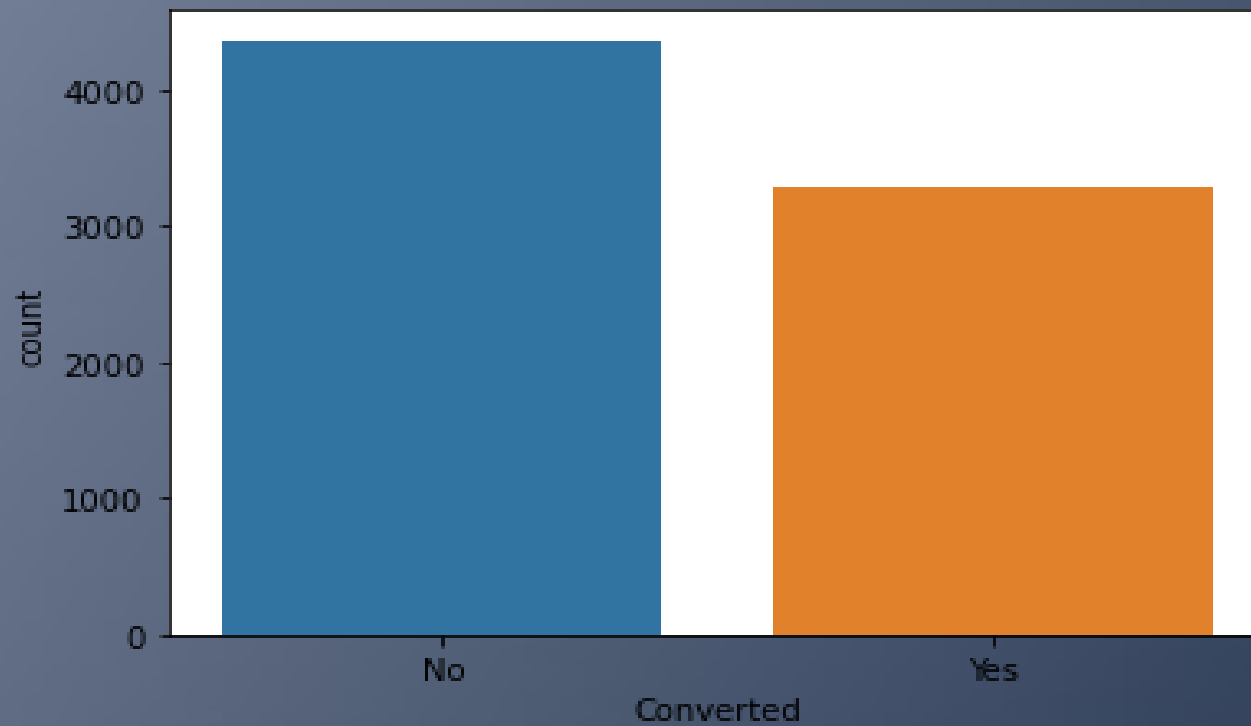
# Problem Solving Approach

1. Cleaning of data, standardizing and inserting missing values in data.
2. Performing Exploratory Data Analysis.
3. Creating dummy variables and scaling of variables.
4. Model Building.
5. Evaluating/testing model accuracy on training dataset.
6. Evaluating/testing model accuracy on test dataset.

# Data Preparation

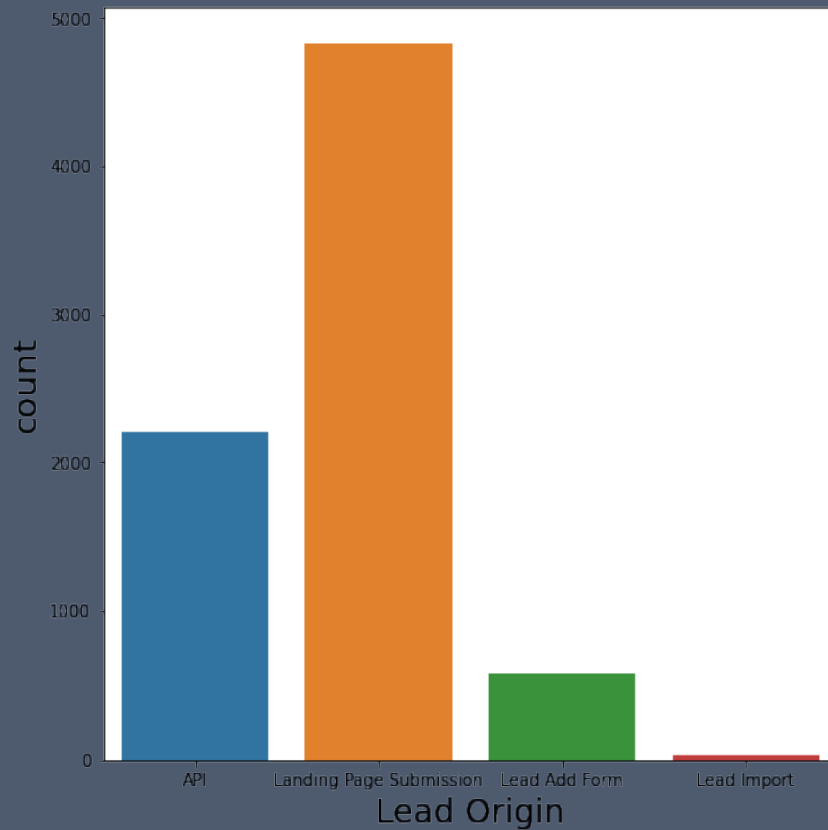
- ❖ Converting the variable values Yes/No into 1/0.
- ❖ Replacing the SELECT value to null value.
- ❖ Dropping those columns having more than 50% missing values.
- ❖ Dropping the rows who have null values less than 3%.
- ❖ Dropping Unnecessary/Unwanted columns.

# Exploratory Data Analysis.

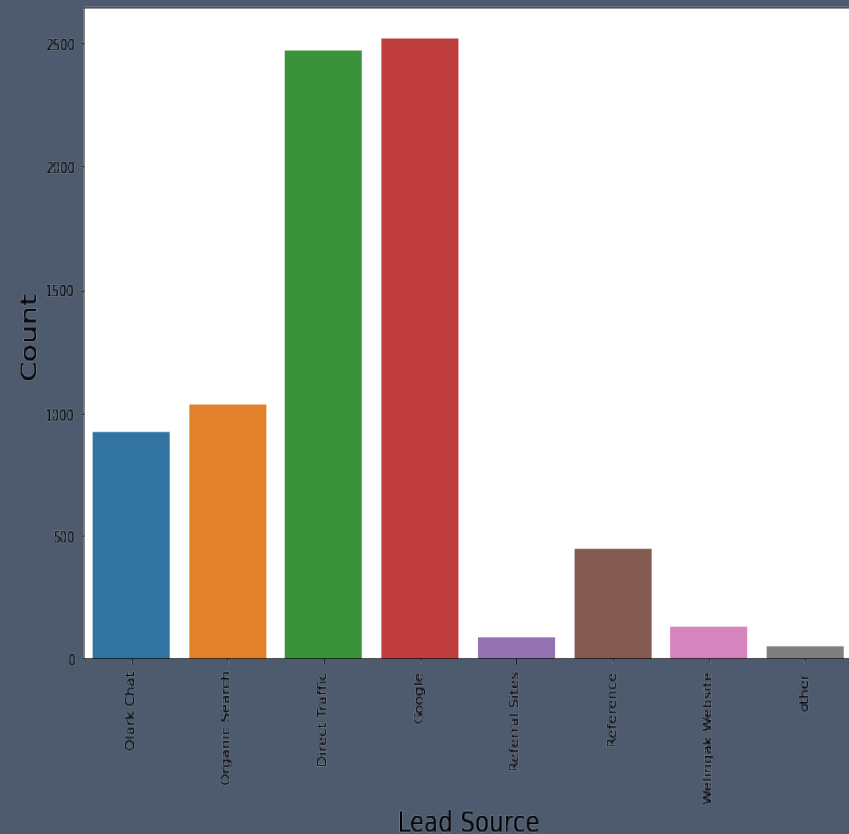


The lead conversion rate before model building is 42%.

# Univariate Analysis

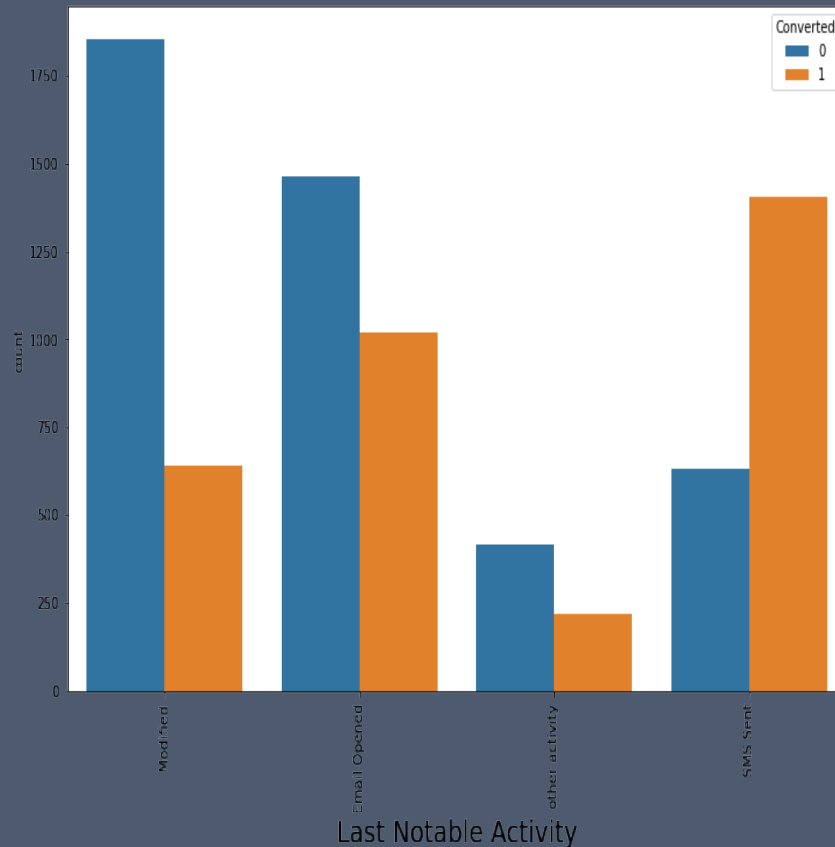


❖ The main lead originates from landing page submission.

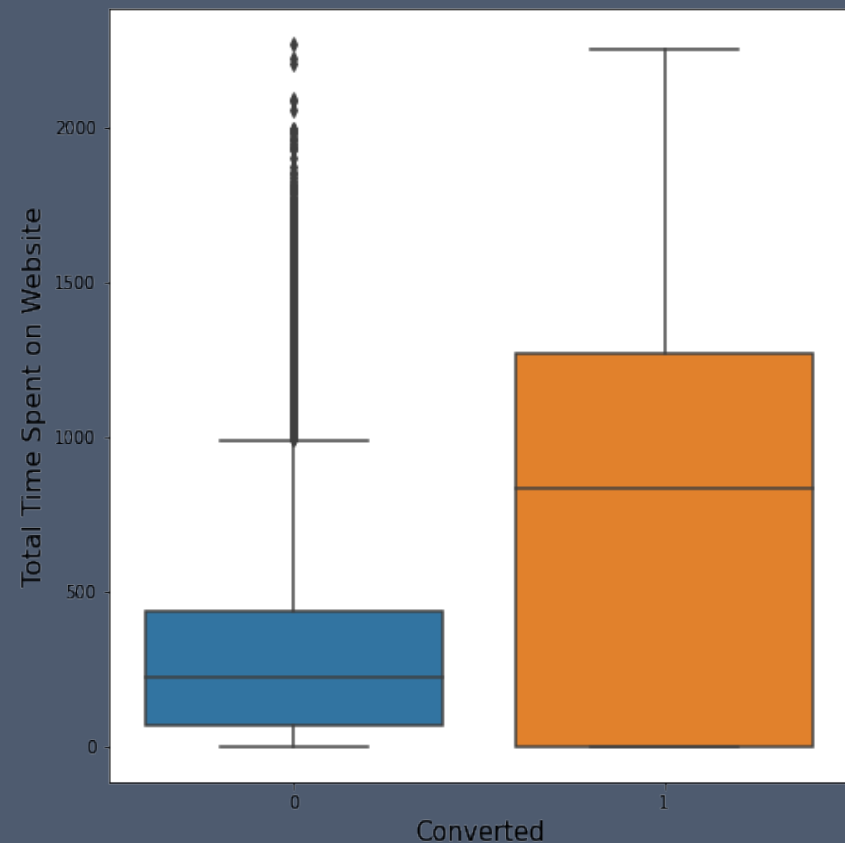


❖ The major lead source is Google and Direct Traffic.

# Bivarient Analysis

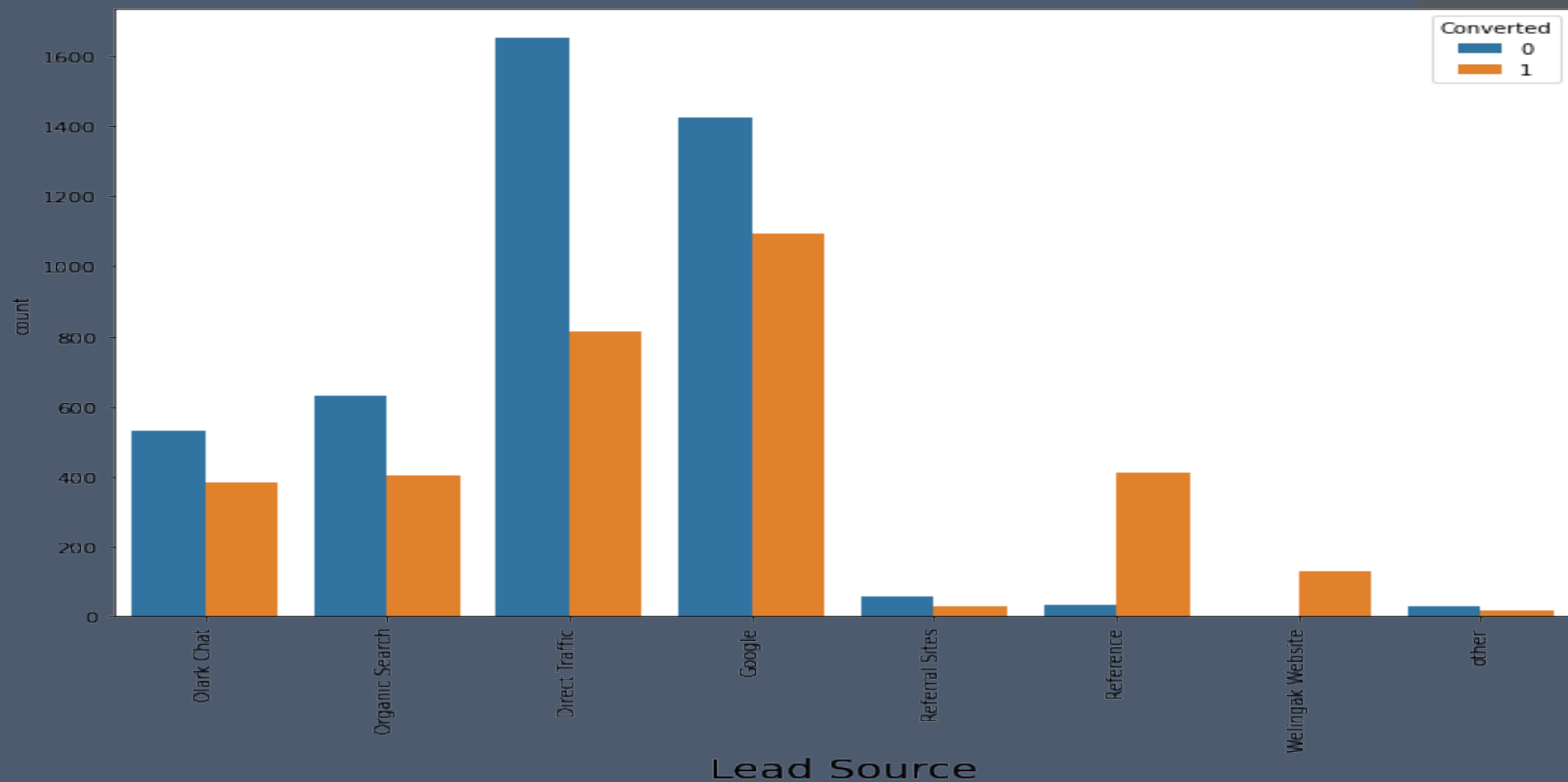


❖ Majority of the customers are converted when last Notable activity was 'SMS Sent'.



❖ As the time spend by the customer increases, their conversion chance also increases.





❖ The Conversion rate for the customer was high when the Lead Source was 'Reference' and 'Welingak Website'

# MODEL BUILDING

STEP 1:

- Dividing the data into dependent (X) and independent variable (y).

STEP 2

- Splitting the data into train-test data in ratio 70:30.

STEP 3

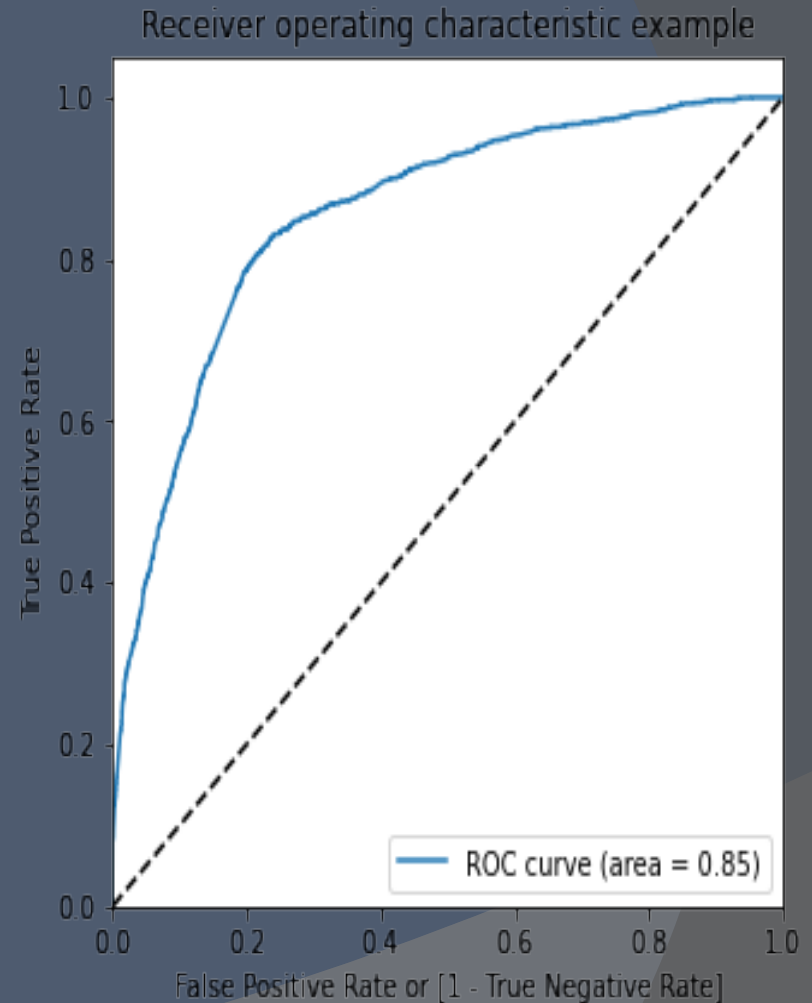
- With the use of RFE selecting the top 15 variables.

STEP 4

- With manual approach selecting those variables having p-values and V.I.F less than 0.05 and 5.

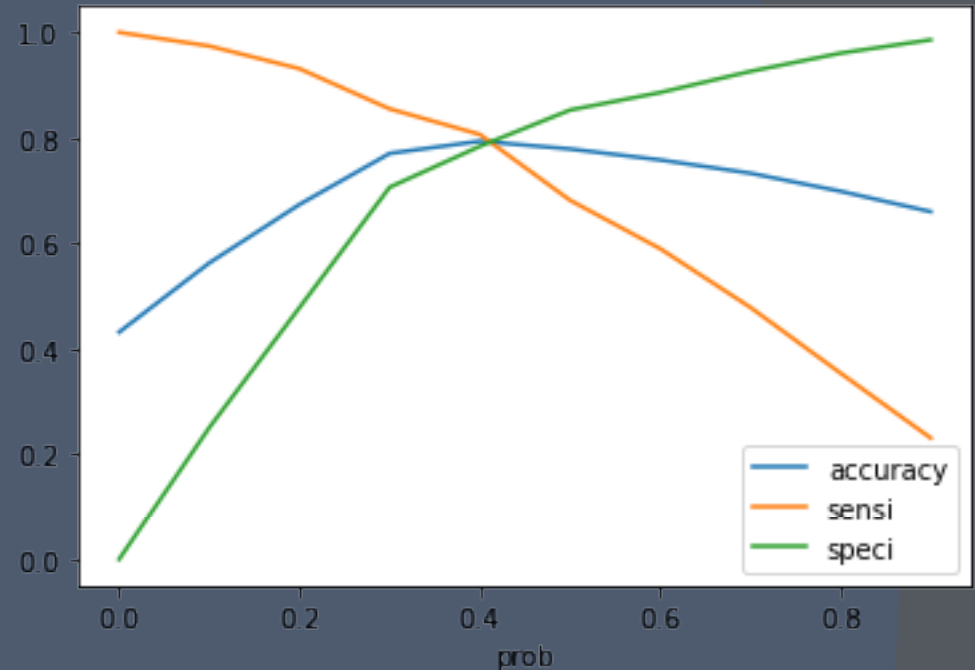
STEP 5

- Predicting the dependent variable.



# MODEL EVALUATION

- ❖ Calculated Accuracy, Sensitivity, Specificity on various probabilities ranges from 0.1 to 0.9.
- ❖ The optimal threshold point comes out to be 0.4 of probability where the customer is predicted as lead whose probability is equal to or more than 0.4.
- ❖ The accuracy of the model comes to 80.64 %.



Accuracy	79.36%
Sensitivity	80.64%
Specificity	78.39%

		PREDICTED	
		NO	YES
ACTUAL	NO	2593	448
	YES	735	1574

## Analysis & Recommendations

- The sales team of the X-Education should focus on the leads having lead origin - lead add form , occupation - Working Professional , Lead source - Wellingak website.
- Form should mandatorily capture important variables like specialization & occupation which will definitely add value to build better model
- We have high recall score than precision score. Hence this model has an ability to adjust with the company's requirements in coming future.
- High Sensitivity will ensure that almost all leads who are likely to Convert are correctly predicted where as high Specificity will ensure that leads that are on the brink of the probability of getting Converted or not are not selected.
- It's better to focus least on customers who do not want to be called about the course.

# Conclusion

- The model has used logistic regression for predicting whether the customer will be converted as 'hot lead' or not.
- The most important features which should be focused more are:
  - 1) Total Time Spend on Website.
  - 2) Lead Origin :- Landing page submission.
  - 3) Lead Source :- Reference and Welingak Website
  - 4) Specialization :- Financial Management.
  - 5) Last Activity :- SMS Sent.
  - 6) Current Occupation :- Working Professional.
- These are the major features where the X education company should focus more on for increasing the conversion rate.