LEAD SCORE CASE STUDY

PRESENTED BY:

MS. NEHA PATIL

MR. RISHIKESH GUNDLA

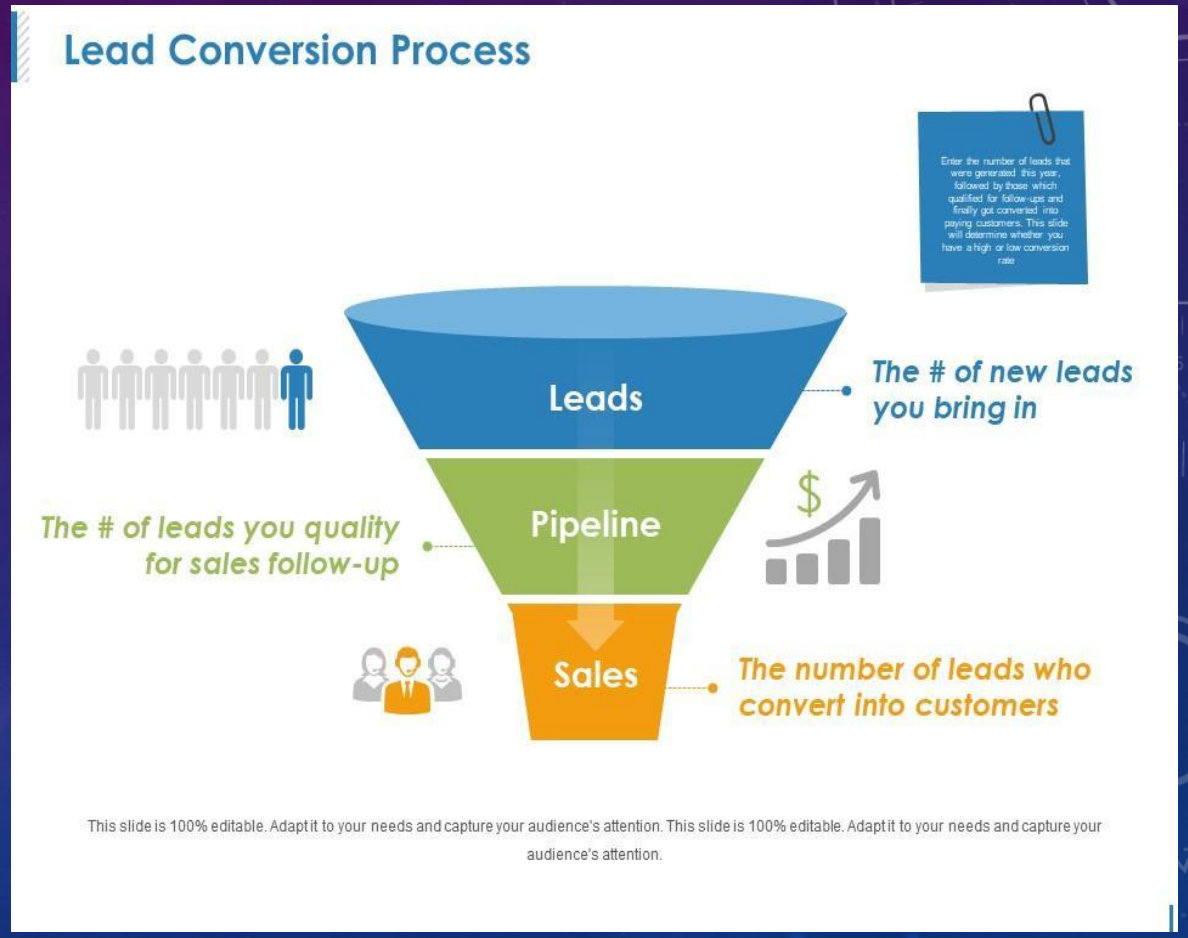
PROBLEM STATEMENT

- 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%.
- Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.
- X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. **The company requires you to build a model wherein you need to assign a lead score 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%.**

LEAD CONVERSION PROCESS

- As you can see, there are a lot of leads created in the first stage (top), but only a few of them become paying clients in the second step. To acquire a greater lead conversion in the middle stage, it is necessary to nurture the potential leads well (i.e. educating the leads about the product, regularly communicating, etc.)

- The company must create a model in which each lead is assigned a lead score so that customers with higher lead scores have a higher conversion chance and customers with lower lead scores have a lower conversion chance. The CEO, in particular, has stated that the objective lead conversion rate should be about 80%.



CASE STUDY OBJECTIVE

- Develop a logistic regression model to assign a lead score between 0 and 100 to each lead, which the organisation may use to target potential leads.

- A greater number indicates that the lead is hot, i.e. favorable to convert, whereas a lower value indicates that the lead is cold and unlikely to convert.

SOLUTION METHODOLOGY

Data Sourcing, Cleaning and Preparation

- Read the Data from the Source
- Convert data into clean format suitable for analysis
- Remove duplicate data
- Outlier Treatment
- Exploratory Data Analysis
- Feature Standardization

Feature Scaling and Splitting Train and Test Sets

- Feature Scaling of Numeric data
- Splitting data into train and test set.

Model Building

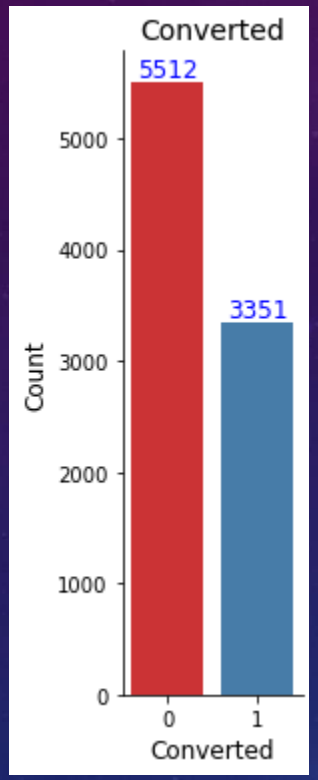
- Feature Selection using RFE.
- Determine the optimal model using Logistic Regression
- Calculate various metrics like accuracy, sensitivity, specificity, precision, recall and evaluate the model

Result

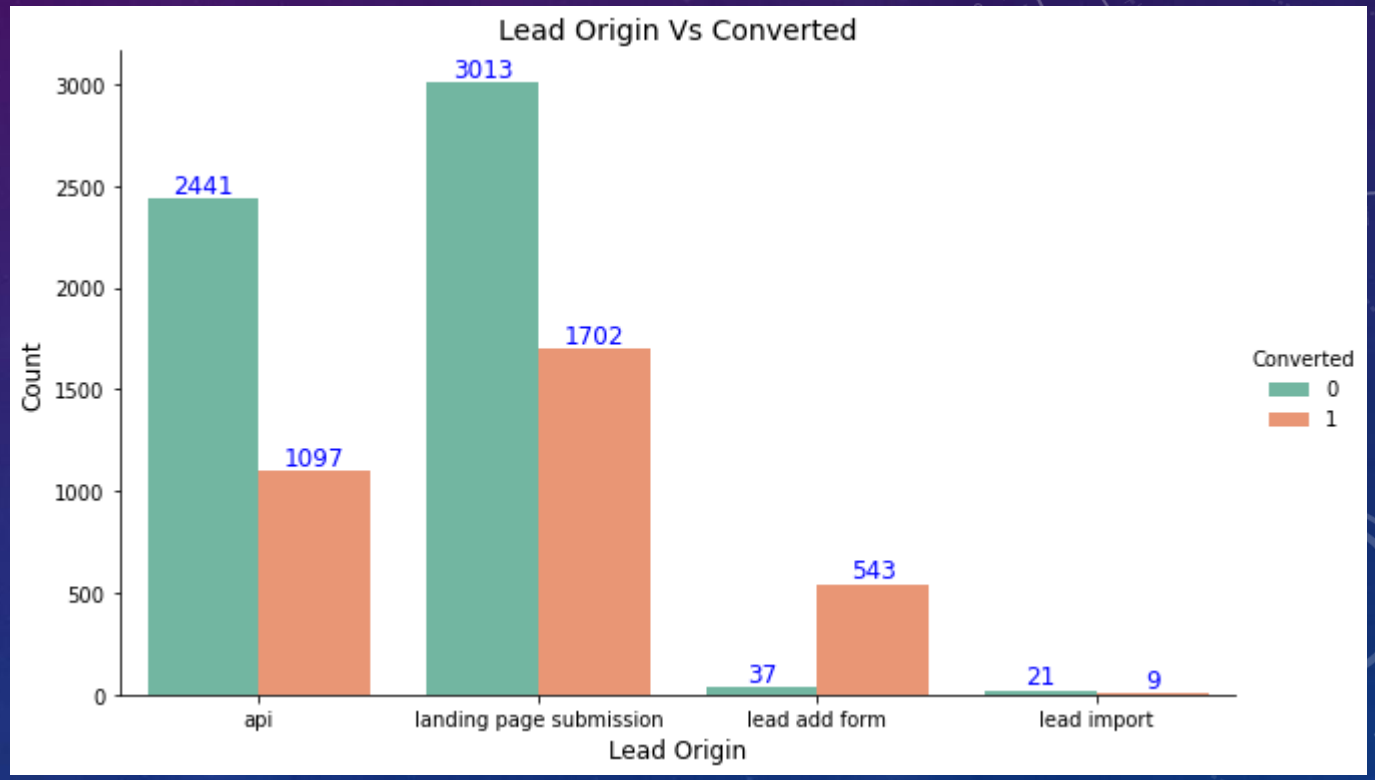
- Determine the lead score and check if target final predictions amounts to 80% conversion rate.
- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics.

EXPLORATORY DATA ANALYSIS

The overall conversion rate is 39%

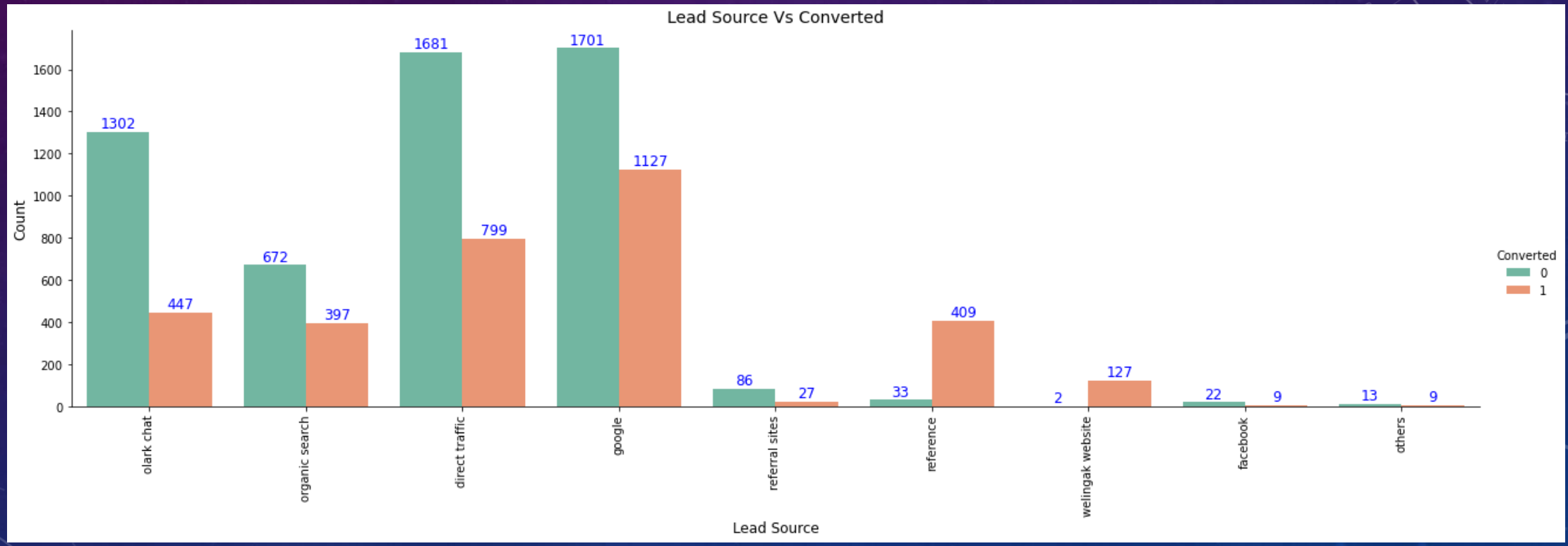


Landing page got the maximum conversion leads followed by API .
This can be a good source to increase the leads in the future.



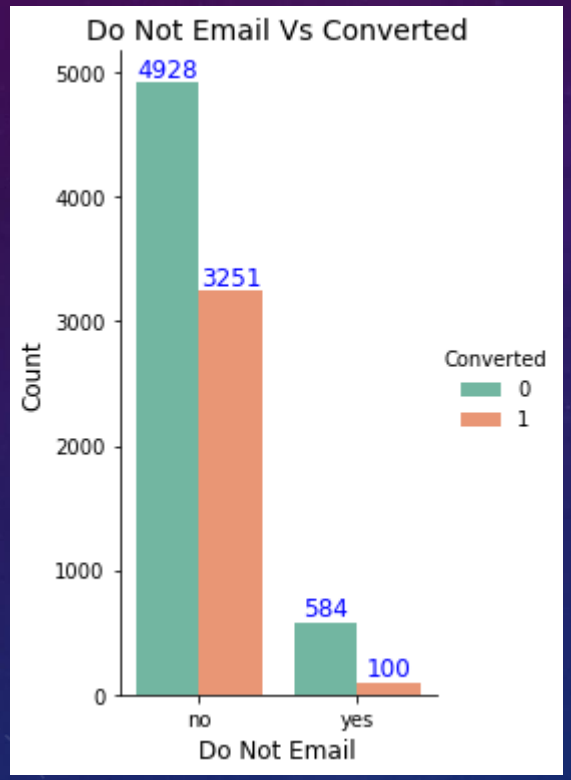
EXPLORATORY DATA ANALYSIS

Most of the conversion happened from Google followed by direct traffic. The source dark chat also has a conversion rate of about 25%.

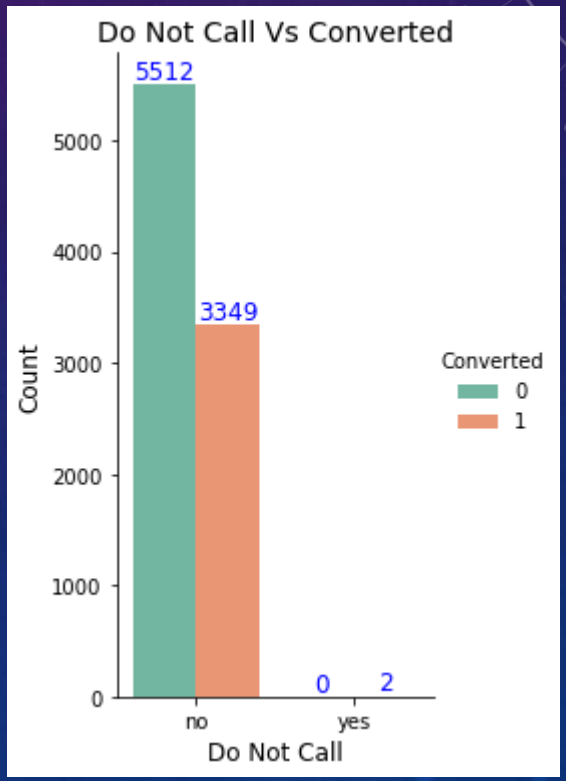


EXPLORATORY DATA ANALYSIS

People who opted out of email subscription have a higher conversion rate than the one who opted in.

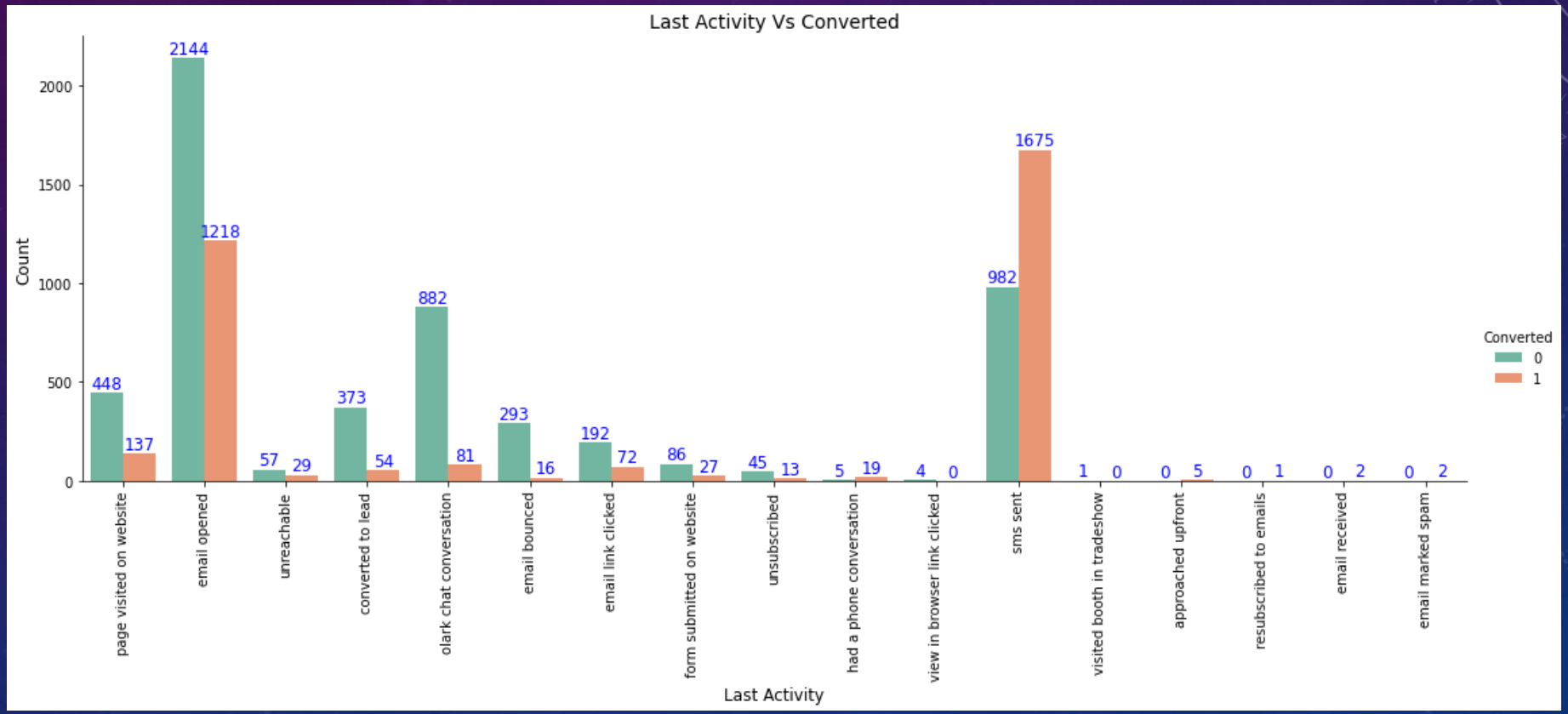


People who opted out of calling option have a higher conversion rate than the one who opted in.



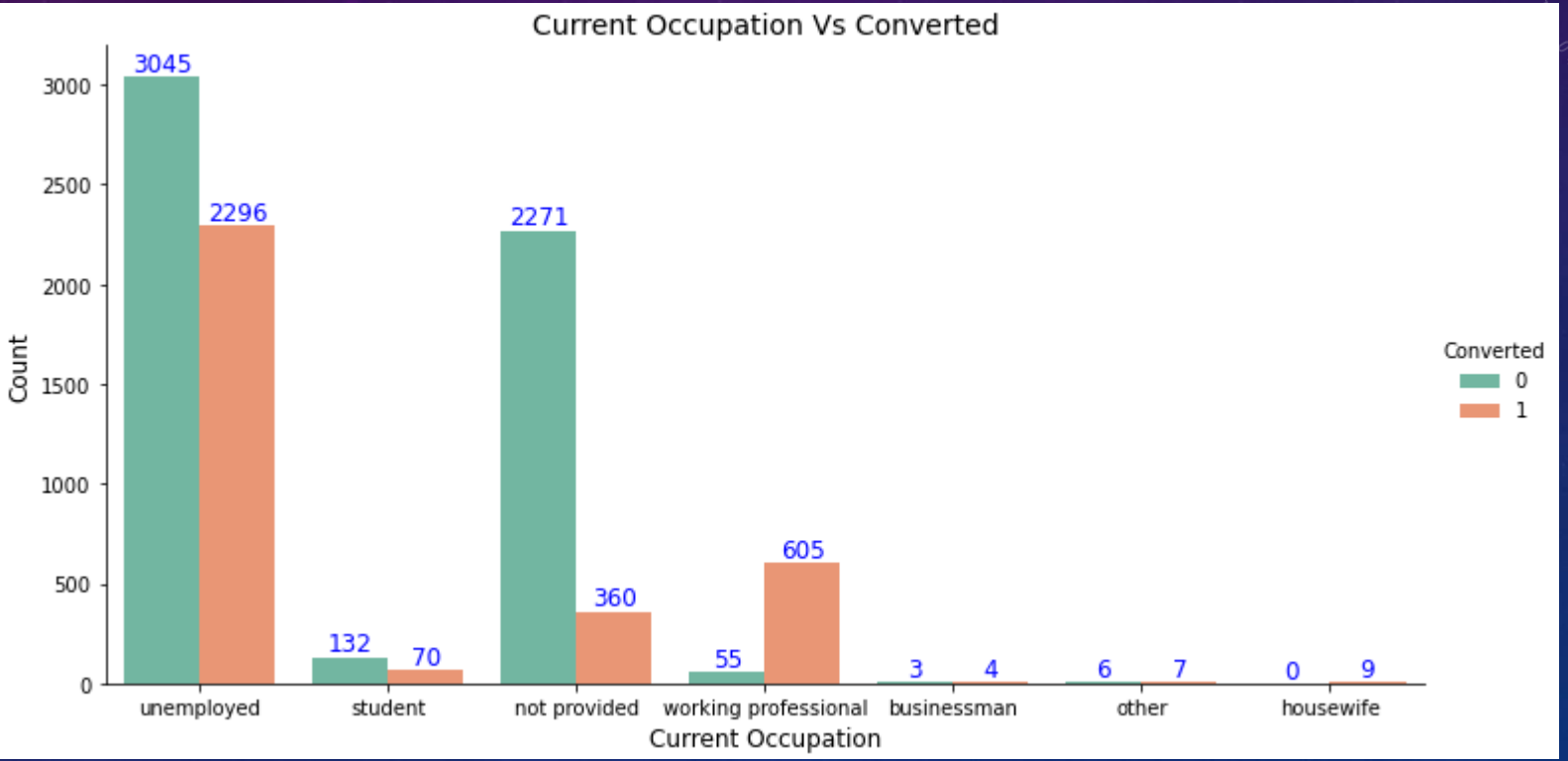
EXPLORATORY DATA ANALYSIS

The activities Email Opened and SMS sent have a very high conversion rate as compared to other activities.



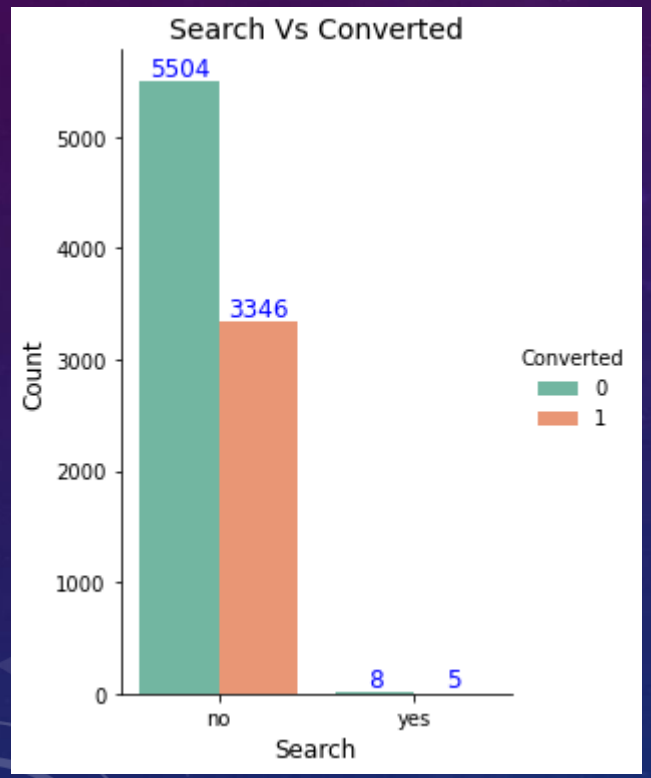
EXPLORATORY DATA ANALYSIS

The Working professionals have a very high conversion rate, this can be a targeted group for future conversions.

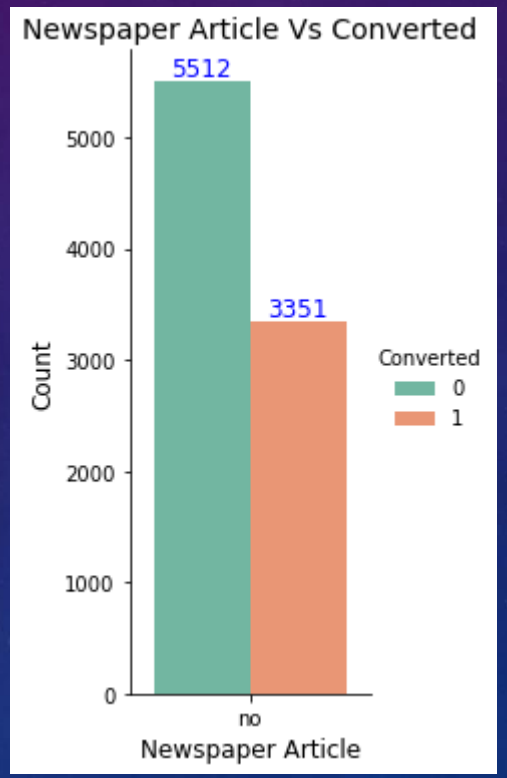


EXPLORATORY DATA ANALYSIS

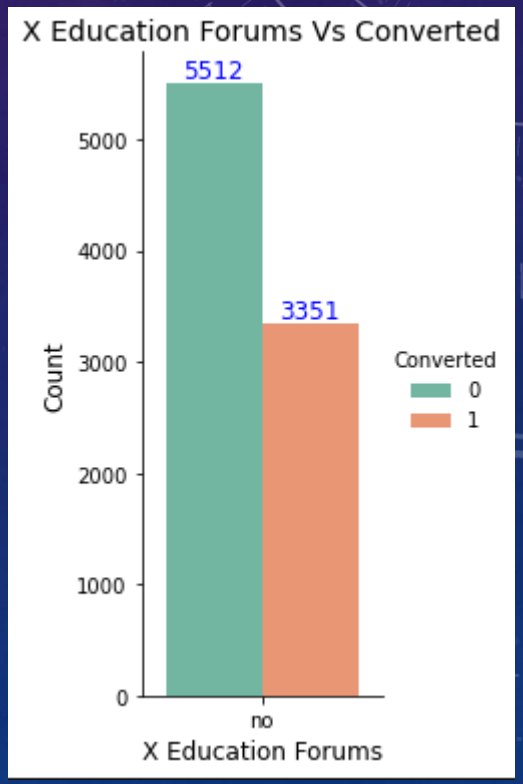
Search is also a very good source for higher conversions, the company can focus upon promotional ads to increase the conversion rates.



Newspaper Article has only one value for all rows - "No" , it is safe to drop this column

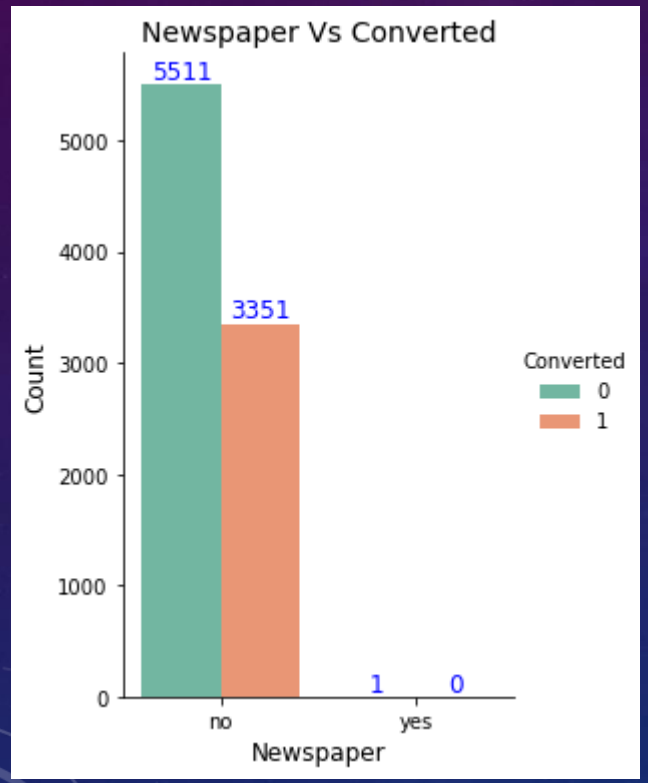


X Education Forum has only one value for all rows - "No" , it is safe to drop this column

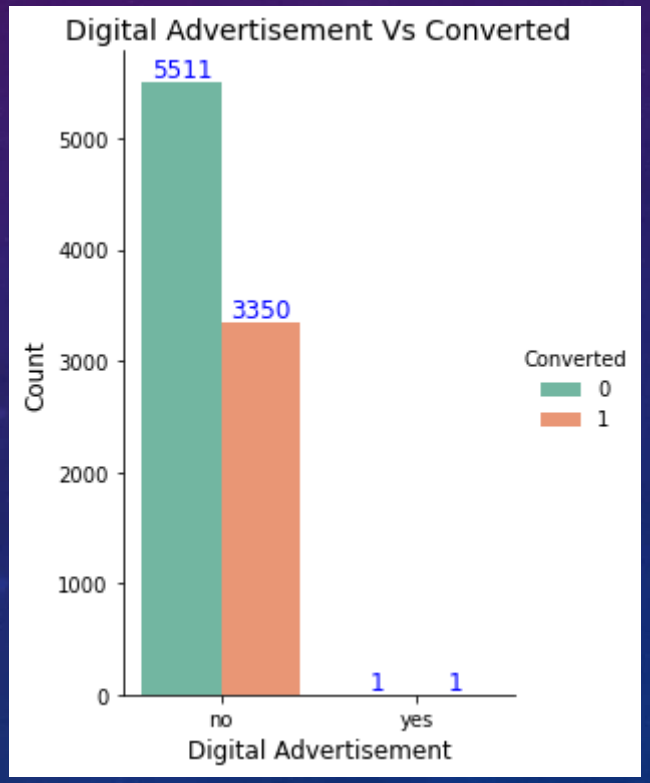


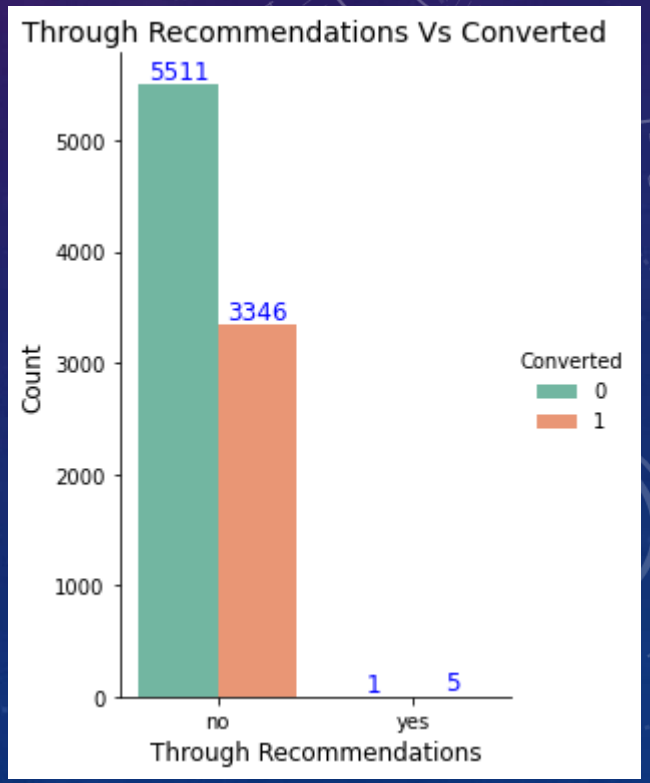
EXPLORATORY DATA ANALYSIS

Since Newspaper column has only one row with "Yes" as the value and further since this lead did not get converted and rest of all the values are "No", we can safely drop the column



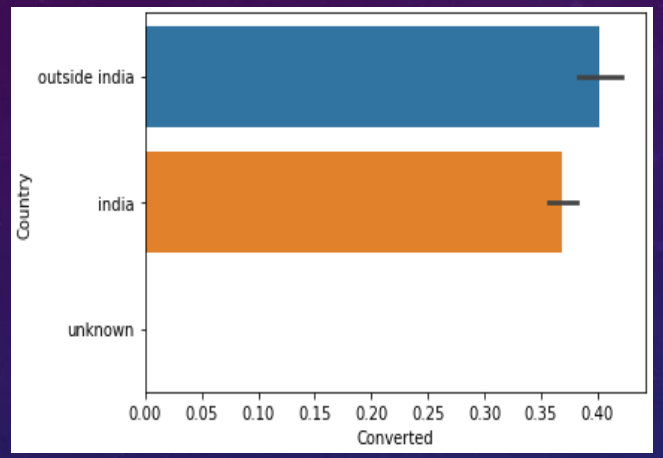
Not much impact was seen on conversion rates through Digital Advertisement and Through Recommendations



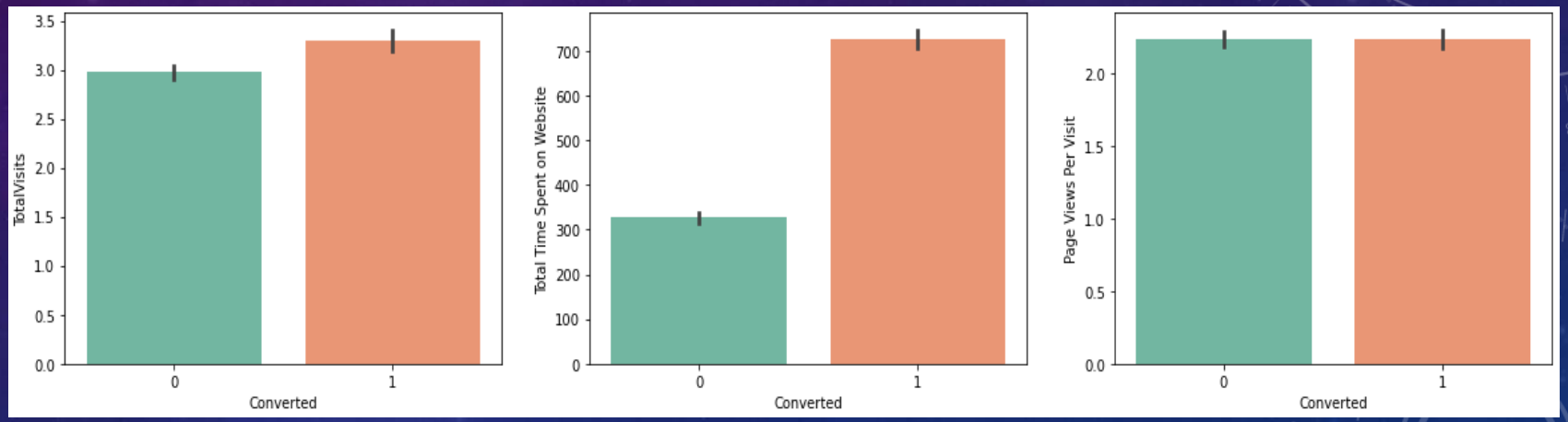


EXPLORATORY DATA ANALYSIS

It is interesting to note that the conversion rate outside India is also very good.

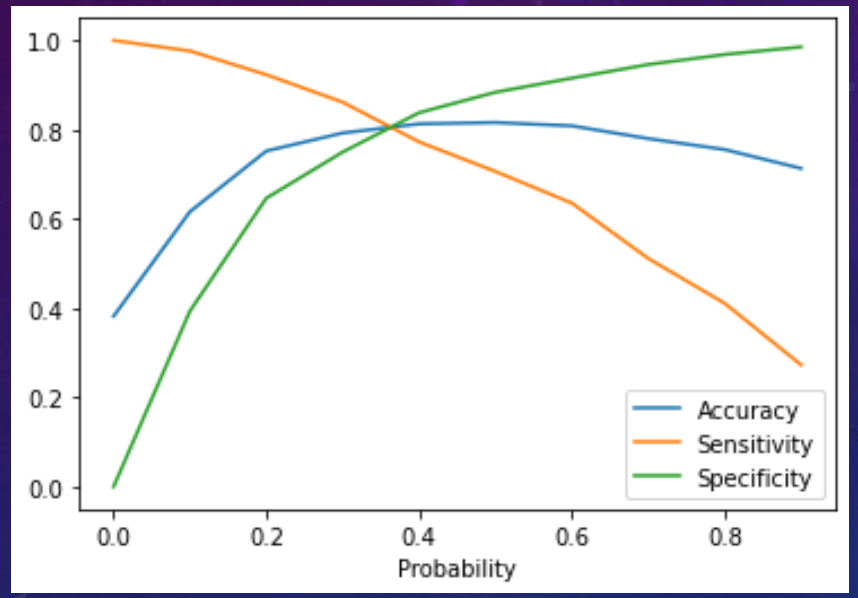


It can be seen that more time the visitors spend on the website, higher the conversion chances are. The company can develop the website in such a way that the visitors spend time to browse more about the courses.



MODEL EVALUATION - SENSITIVITY AND SPECIFICITY ON TRAIN DATA SET

The graph depicts an optimal cut off of 0.35 based on Accuracy, Sensitivity and Specificity



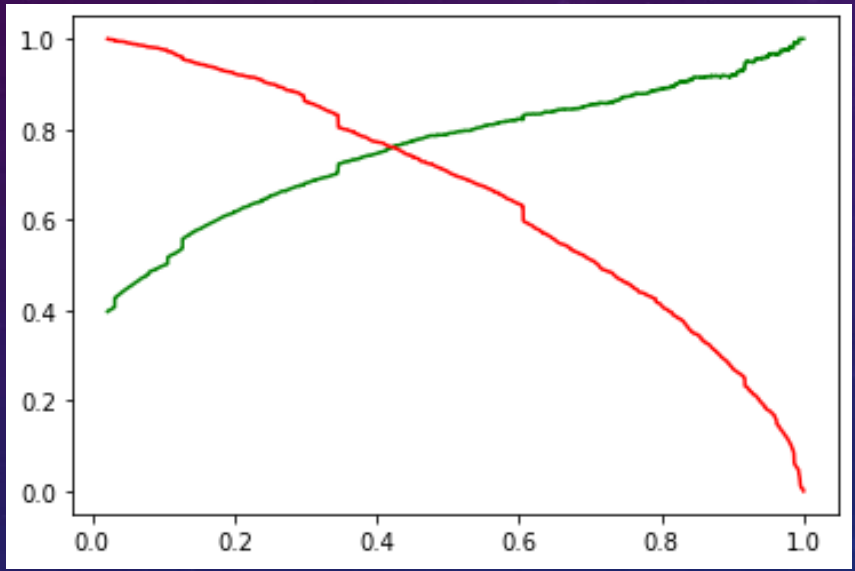
Confusion Matrix

3112	722
468	1902

Accuracy - 81%
Sensitivity - 80 %
Specificity - 81 %

MODEL EVALUATION - PRECISION AND RECALL ON TRAIN DATASET

The graph depicts an optimal cut off of 0.41 based on Precision and Recall



Confusion Matrix

$$\begin{pmatrix} 3231 & 603 \\ 548 & 1822 \end{pmatrix}$$

Precision - 75 %
Recall - 76 %

MODEL EVALUATION - SENSITIVITY AND SPECIFICITY ON TEST DATASET

Confusion Matrix

$$\begin{pmatrix} 1426 & 252 \\ 248 & 733 \end{pmatrix}$$

Accuracy - 81 %
Sensitivity - 79 %
Specificity - 81 %

CONCLUSION

- **Accuracy, Sensitivity and Specificity** values of test set are around **81%, 80% and 81%** which are approximately closer to the respective values calculated using trained set.

- Also the lead score calculated in the trained set of data shows the conversion rate on the **final predicted model is around 80%. Hence overall this model seems to be good**

- Also, after analyzing, the following variables(Top 10) were found to be mostly important for getting potential buyers converted:

  - **TotalVisits** : Total number of people who visited the profile
  - **What is your current occupation_housewife** : Visitors who belong to the housewife category
  - **Last Activity_email marked spam (from Last Activity)**
  - **Last Activity_email received (from Last Activity)**
  - **Lead Source_welingak_website (from Lead Source)**
  - **Total Time spent on the website**
  - **Lead Source_reference (from Lead Source)**