AGENDA

- LEAD SCORE – Problem statement
- Assumptions
- Approach
- Recommendations based on Analysis

PROBLEM STATEMENT

- An education company named X Education sells online courses to industry professionals. The Marketing is done on various platforms such as google, Facebook etc. Leads are generated when the customers visit the website or through referrals. Once these Leads are generated, sales team starts approaches to make them as converted leads through various business strategies such as dropping mails, SMS etc.
- X Education gets a lot of leads, its lead conversion rate is very poor. This case study identifies the Leads as hot leads and cold leads based on their Lead score. This helps the education company to target particular hot leads to increase their conversion rate.

ASSUMPTIONS

- Based on the Marketing and sales perspective, its good to have few assumptions before going in-depth analysis with the dataset.

  a. People who are working professionals tend to buy the course to improve their skillsets and positions in their current role.

  b. Majority of Leads have not mentioned their specializations. The reason behind this could be they are either a student or they haven't identified their specialization yet

  c. People who tends to spend more time on website, may be tend to buy the course.

APPROACH ON PROBLEM STATEMENT

a. Understand the business scenario and problem statement.
b. Inspecting the data and export the data.
c. Data cleaning and analysis.
d. Model building
e. Prediction
f. Conclusion/ Insights

APPROACH

a. The Dataset consists of 9240 rows and 37 columns and prospect Id and Lead number tends to be the unique code given to the Lead.

b. Verify the columns, datatypes, key factors that can be utilised for analysis.

b. Converison is the Target variable which decides whether the Leads are converted or not.

c. Outliers are handled for important fields using box plot and quantile percentages.

DATA CLEANING AND ANALYSIS

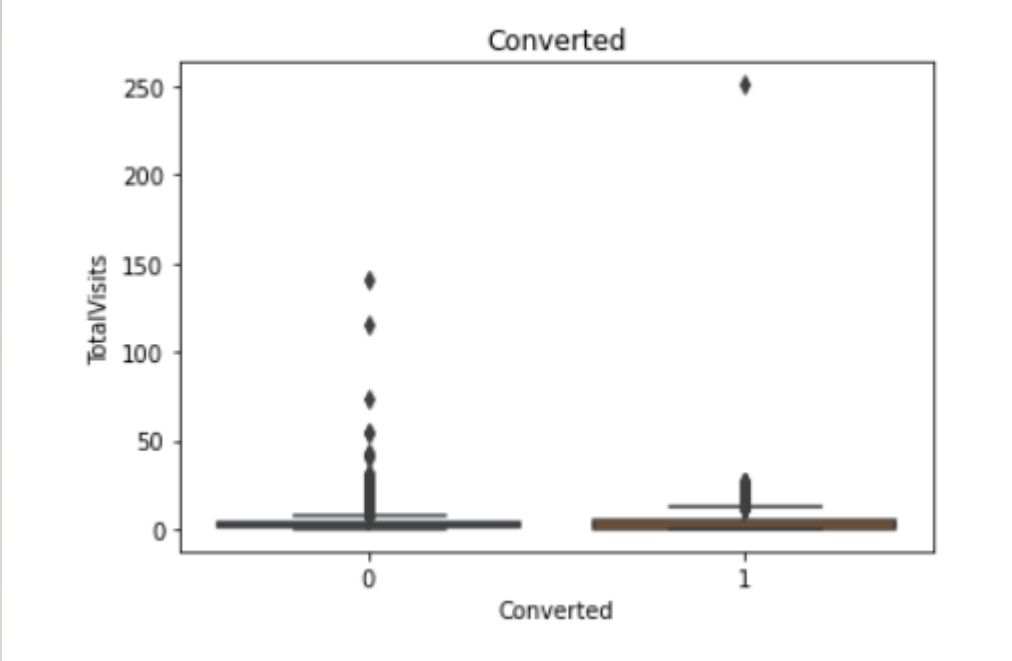
- Treating the missing values:

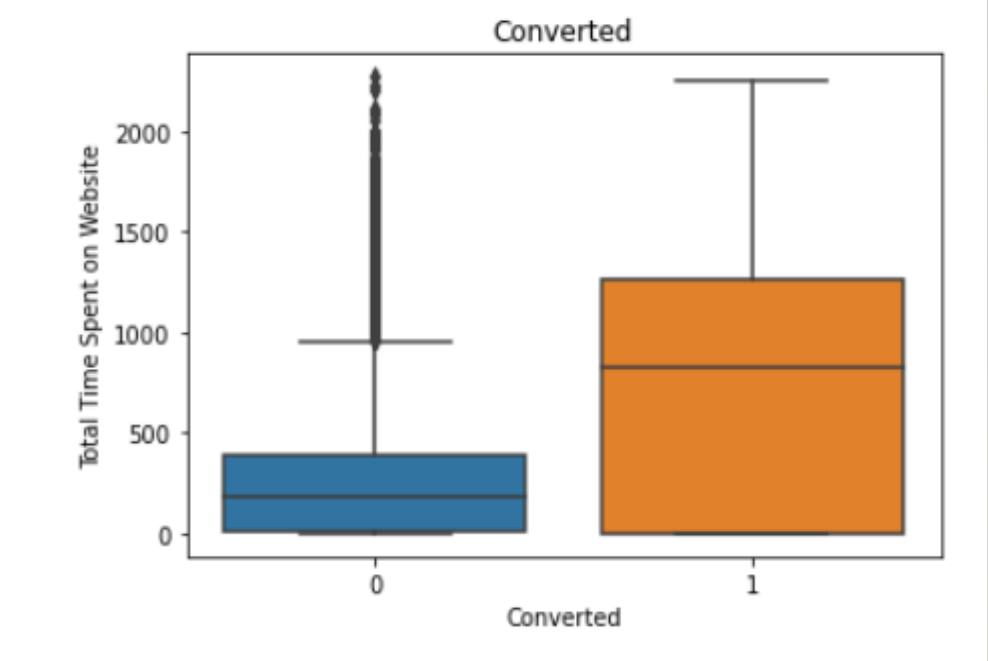
a. Percentage of missing values in the dataset are higher. Hence need to deal carefully in-order to reduce anomalies and wrong assumptions. We could see that there were columns with major missing values like country, Lead profile etc.

All these missing values are treated either by removing or by replacing with mode values or "Not specified". Other than this, the fewer missing values, were dropped as per rows.

There were few columns with "Select". Hence we have imputed it as 'nan'.

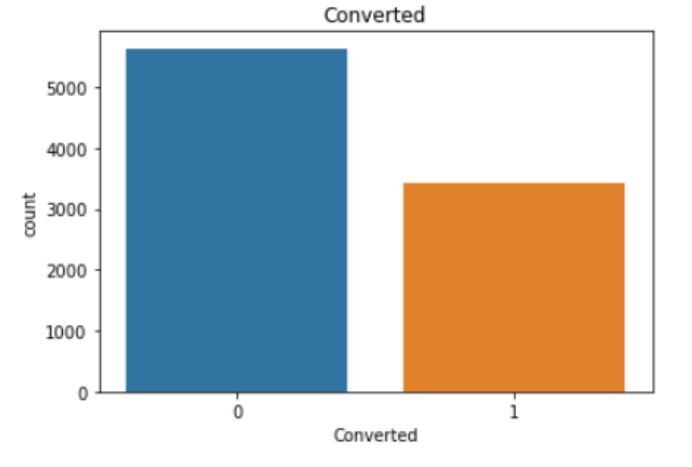
OUTLIERS:

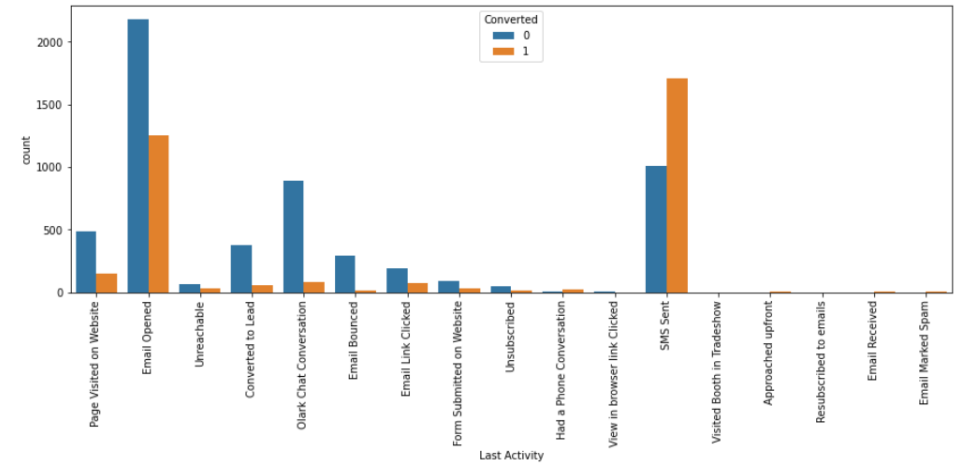

Converted
TotalVisits
250
200
150
100
50
0
0
1
Converted


Converted
Total Time Spent on Website
2000
1500
1000
500
0
0
1
Converted

DATA ANALYSIS:

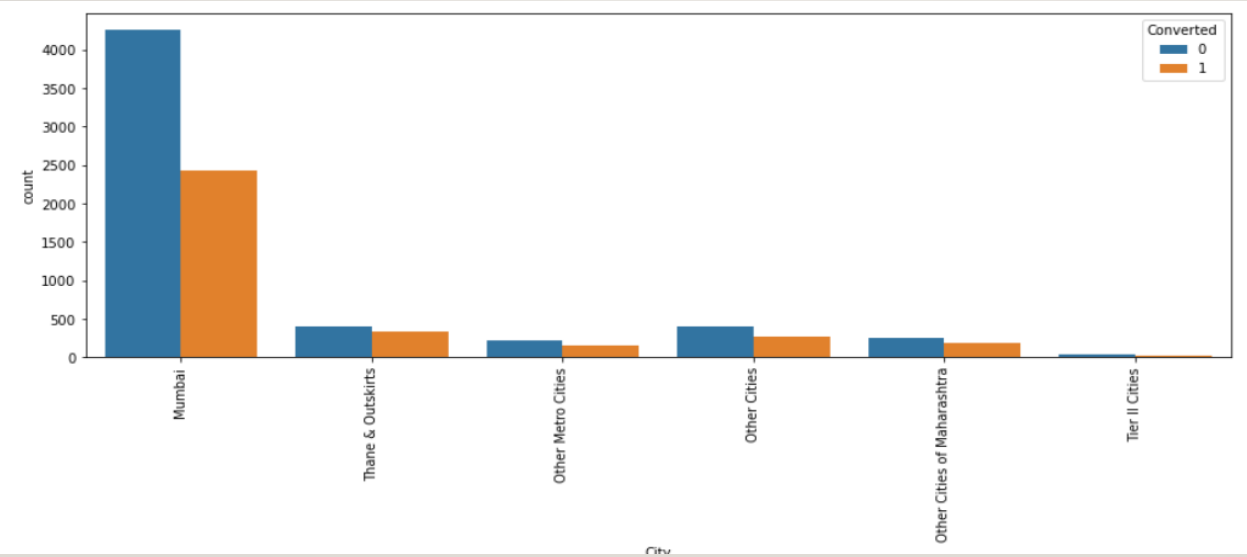
- Conversion rate is very less from the data set (approx. 36%)
- People who are converted to hot leads have last activity majority is through Emails and SMS
- Marketing and sales team need to come up with new ideas to improve conversion rate from other sources included in the last activity.





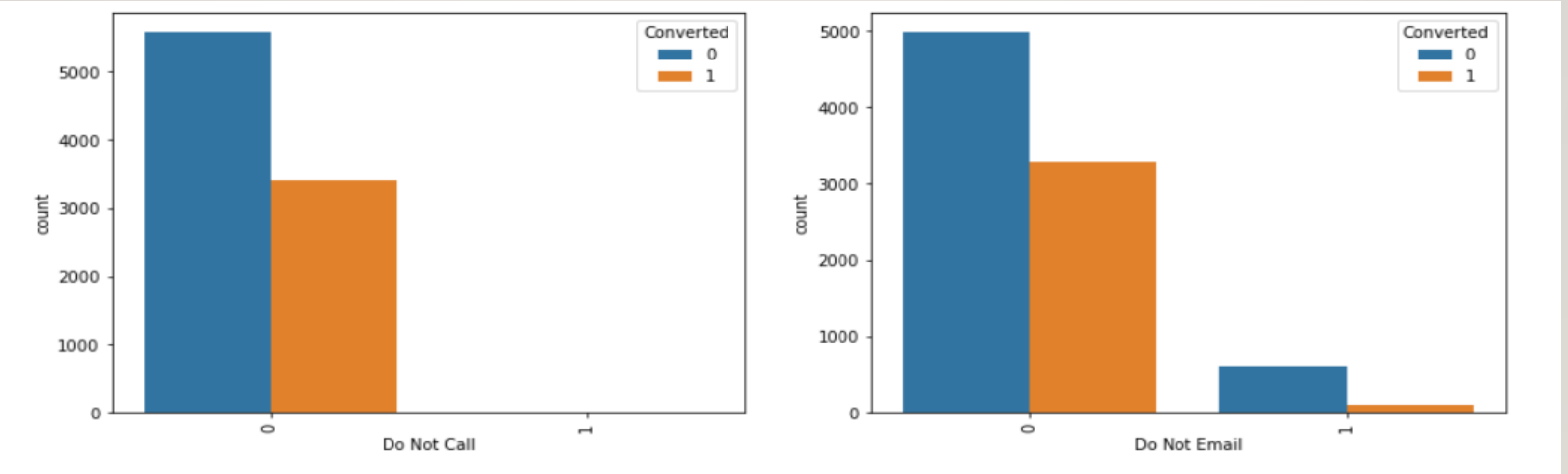
ANALYSIS :

- Majority of leads are generated from Mumbai city in India.
- Lead conversion rate is good in other cities but need to attract to get many leads into their business



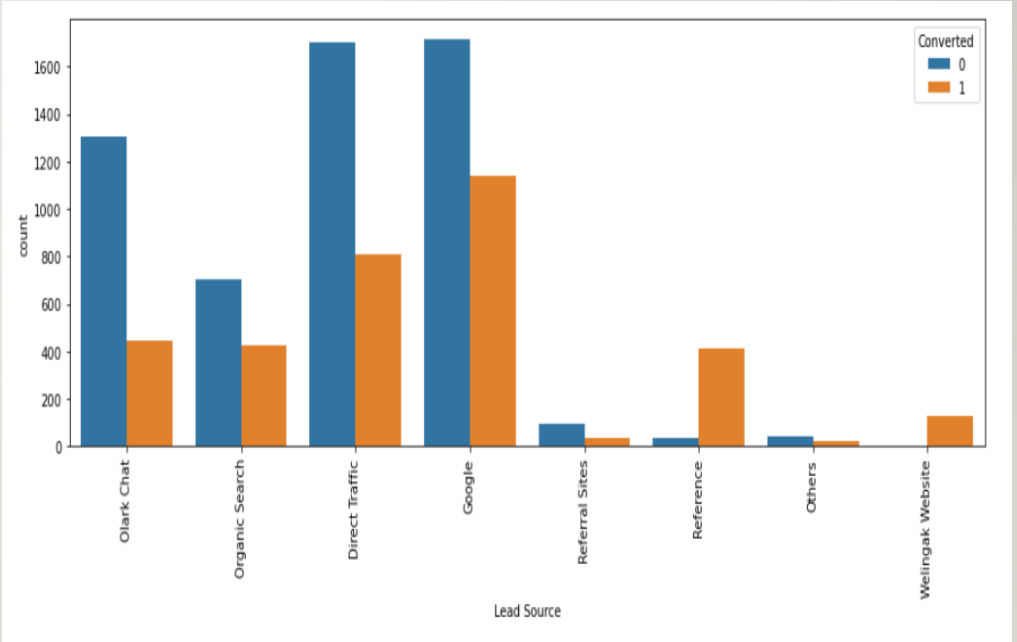
ANALYSIS:

- There is not much inferences to be drawn ‘do not call’ and ‘do not mail’ as major leads are converted based on last activity of mails and SMS.
- However, we need to monitor when is the right time to send SMS and mails without too much frequency.



ANALYSIS:

- Major part of Lead sources are obtained through volunteering and organic way through social sites and google search.
- Conversion rate is quiet high with the help of references. Hence Marketing team needs to concentrate on existing enrollers to help them get them leads.



MODEL BUILDING :

- Model Building started with dummy variable creation for categorical columns and original columns are dropped from the dataset
- There after, the dataset was divided into 80/20 ratio for training and testing.
- Using RFE model, 15 columns were identified and used for prediction.
- Using standard scaler, we have performed scaling on continuous variables.
- We got a accuracy of 92% for the conversion rate based on these columns on testing data.

CONCLUSION:

- Top three variables in the model which contribute most towards the probability of a lead getting converted are :
  a. Total Time Spent on Website
  b. Last_notable_activity_sms_sent
  c. Tags will revert after reading the email.

- Make phone calls to the leads having lead score of greater than 60.
- Make phone calls to the people spending more time on website and are currently working professionals.
- Few Interns should be dedicated to make the website more engaging and user friendly
- Leads with Last activity as SMS sent should be followed up regularly as they have high chance of being converted.