LEAD SCORE CASE STUDY PRESENTATION

Harshavardhana AS
Prakash
Trivikram Chausalkar

Problem Statement

- The X company advertises the online courses for their customers and aims at maximizing the profit.
- The customers who gets converted are considered success by the company , currently the company has only about 30% as lead conversion rate.
- The company needs to increase this conversion rate to atleast 80% to better suit their ROI

PROCEDURE FOLLOWED

1. Loaded the data from ‘Leads.csv’ file and inspected the columns and its descriptive properties and data types.

2. The columns which are of not importance and having missing values >40% are removed.

3. Performed merged bar chart on categorical variables vs the converted (target variable) to understand the feature importance.

4. Identified Top-10 correlation features between the columns and dropped the correlated columns whose value is >=60% to avoid multi-collinearity .

5. Performed binary encoding to two-level categorical variables and one-hot / dummy encoding to other available categorical variables with levels greater than 2.

INSPECTION OF DATA

1. The leads.csv data sheet contains 9240 rows and 37 columns.

2. Each column in the leads.csv data represents the feature information required for model building.

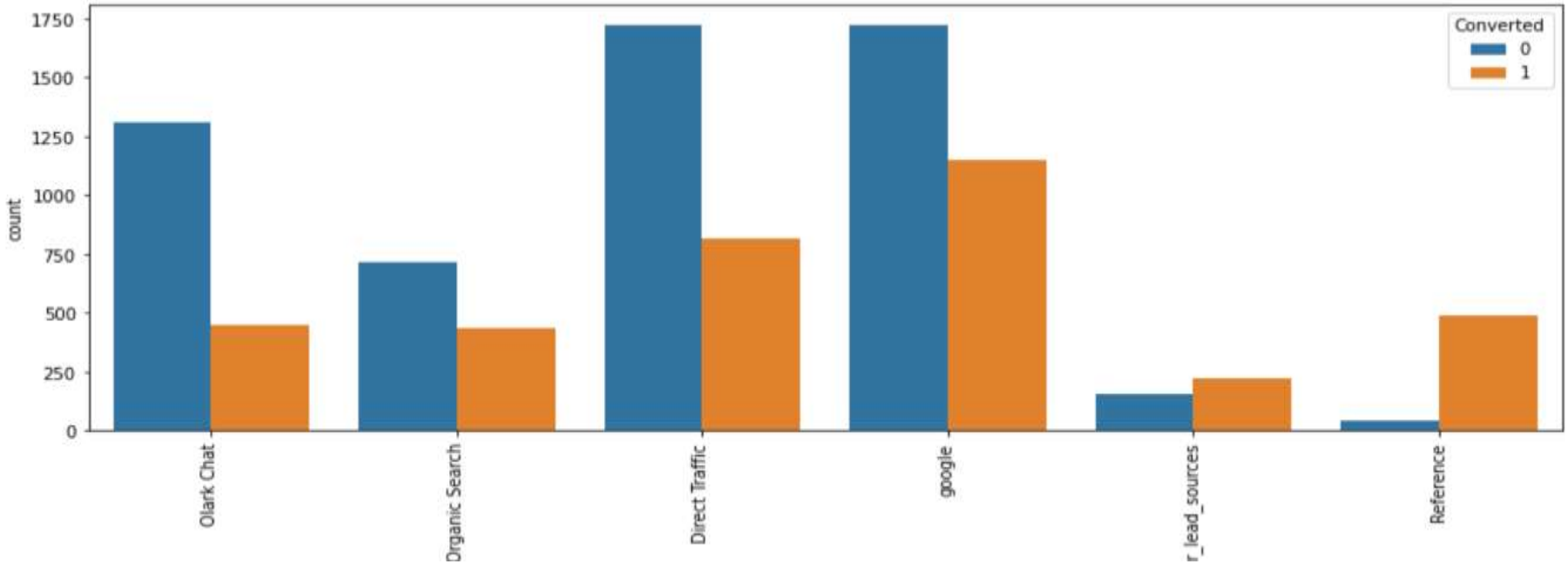
3. The target ‘Converted’ column in leads.csv refers to the conversion of the customers i.,e 0 represents the customers not converted and 1 means converted

4. We have only 6 numerical columns and 31 categorical columns.

5. We identified outliers in the columns ‘Total Visits’ and ‘Page View Per Visit’ and handled accordingly.

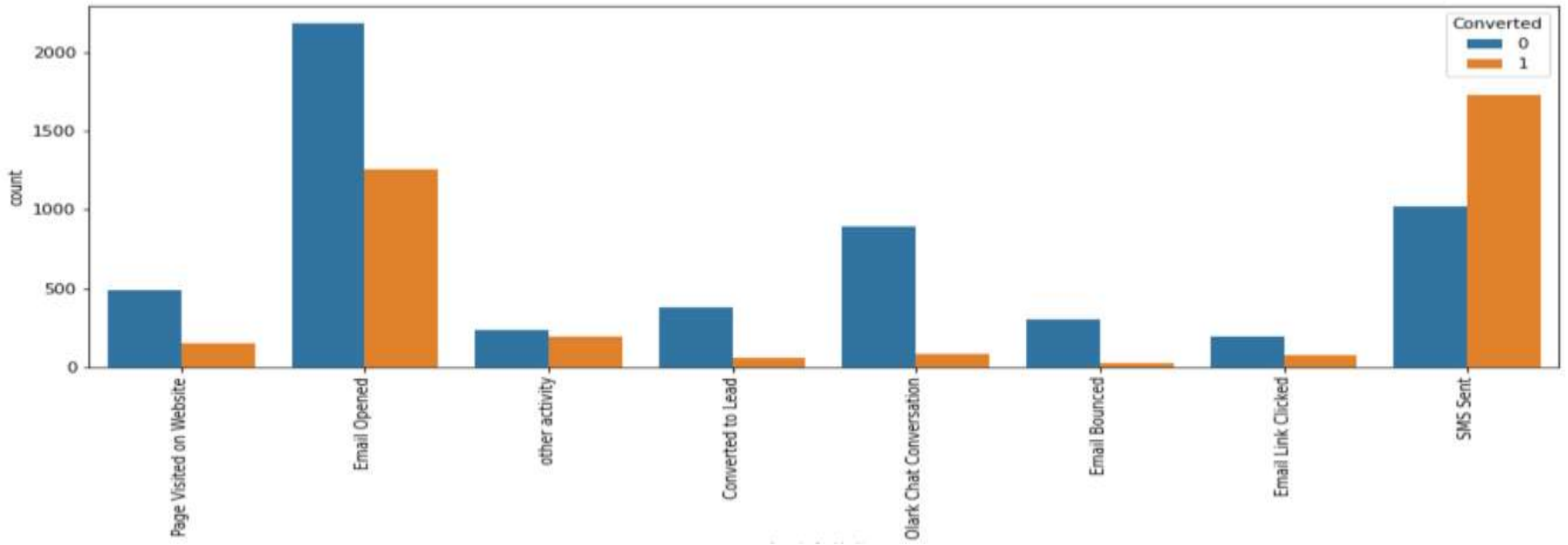
Analyzing Important Categorical Variables

Lead Source vs Converted



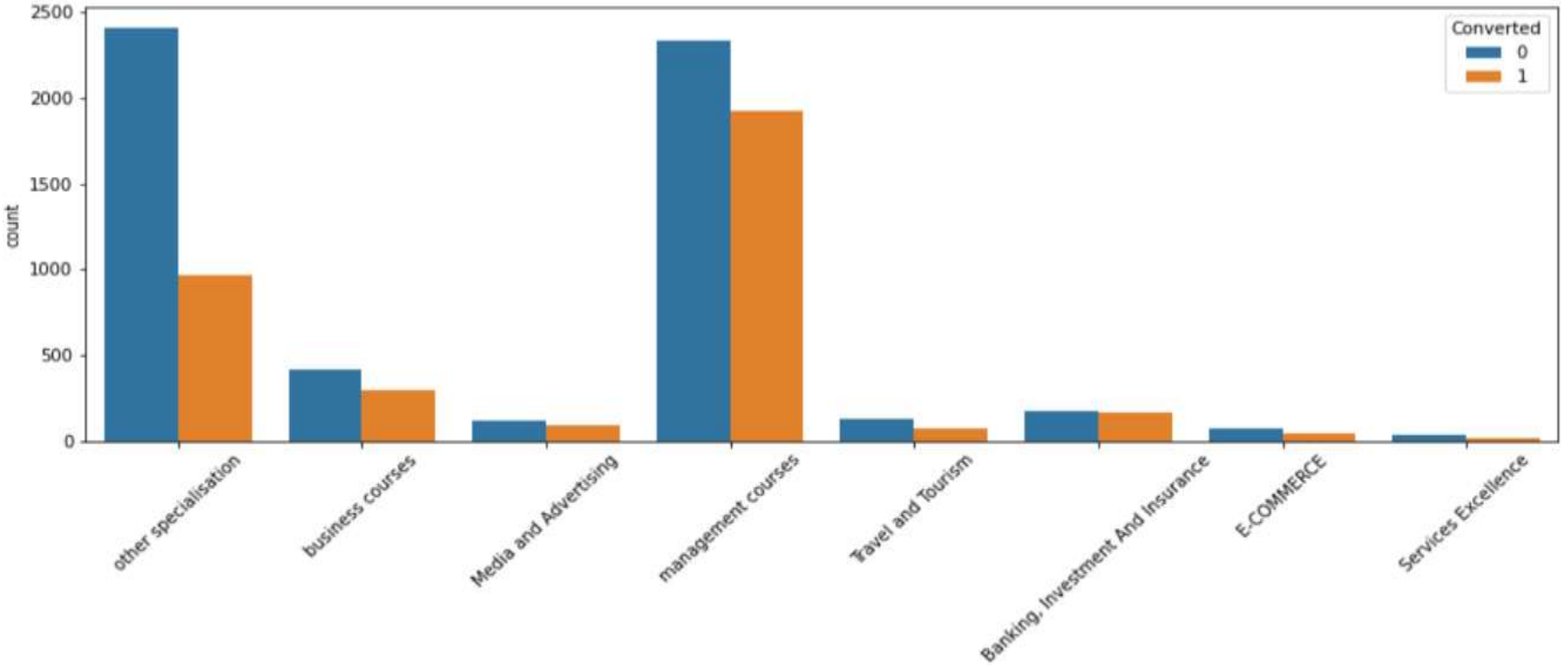
Lead source with Google / Direct Traffic has the highest lead conversion rate

Last Activity vs Converted



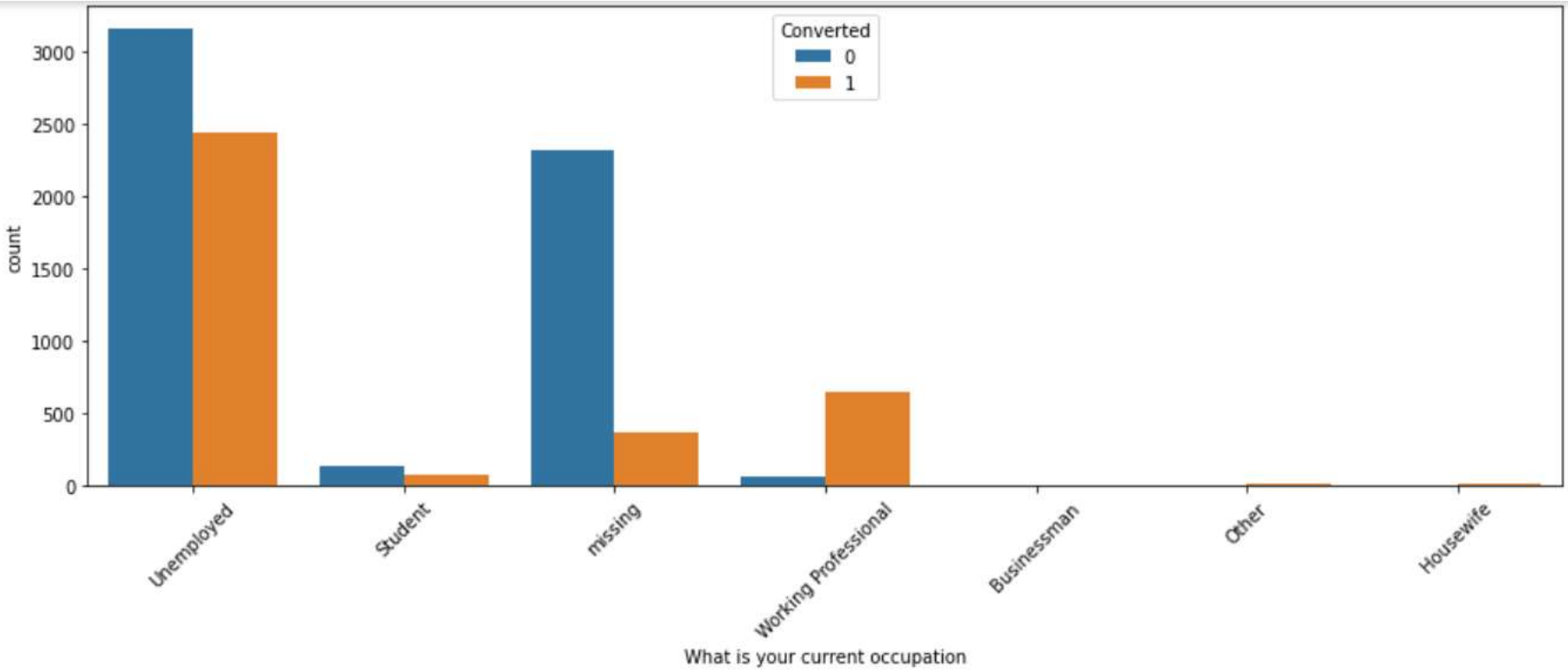
- SMS Sent has greater leads conversion rate proportionally

Specialization vs Converted



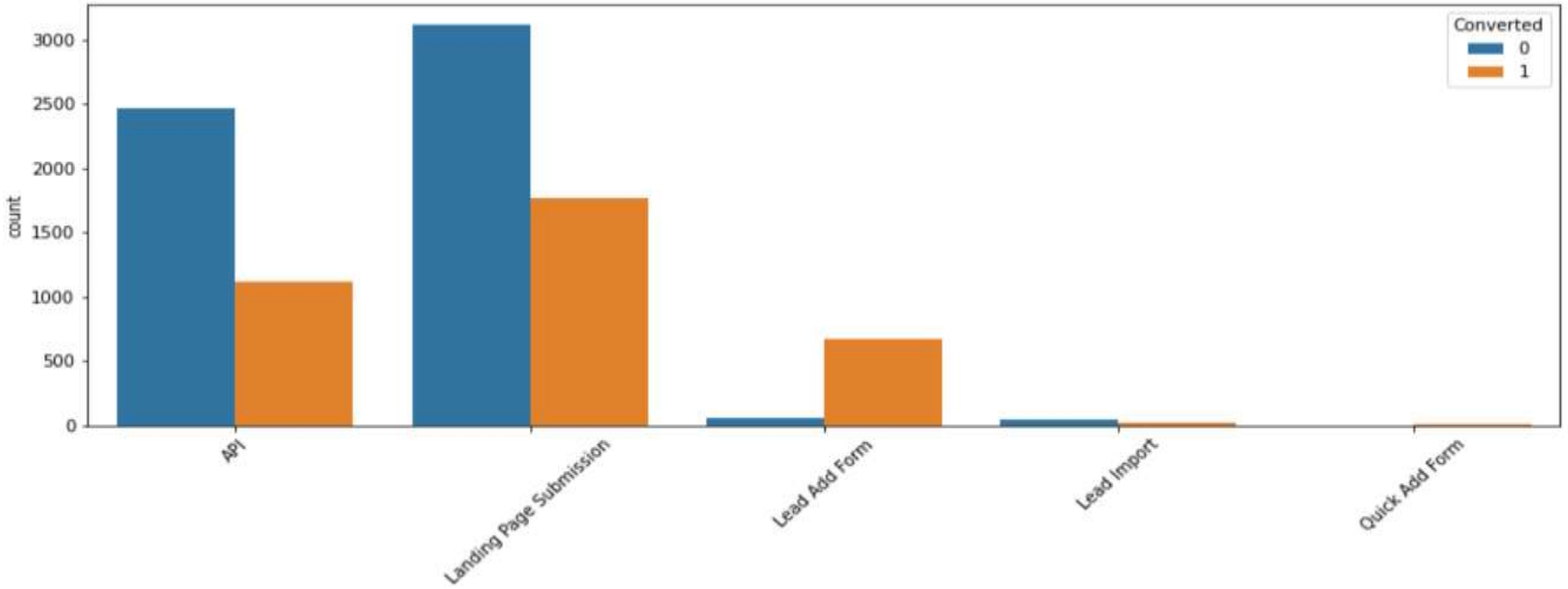
Management type of courses has the highest lead conversion

Occupation vs Converted



Working Professional and unemployed generally gets converted when compared to students

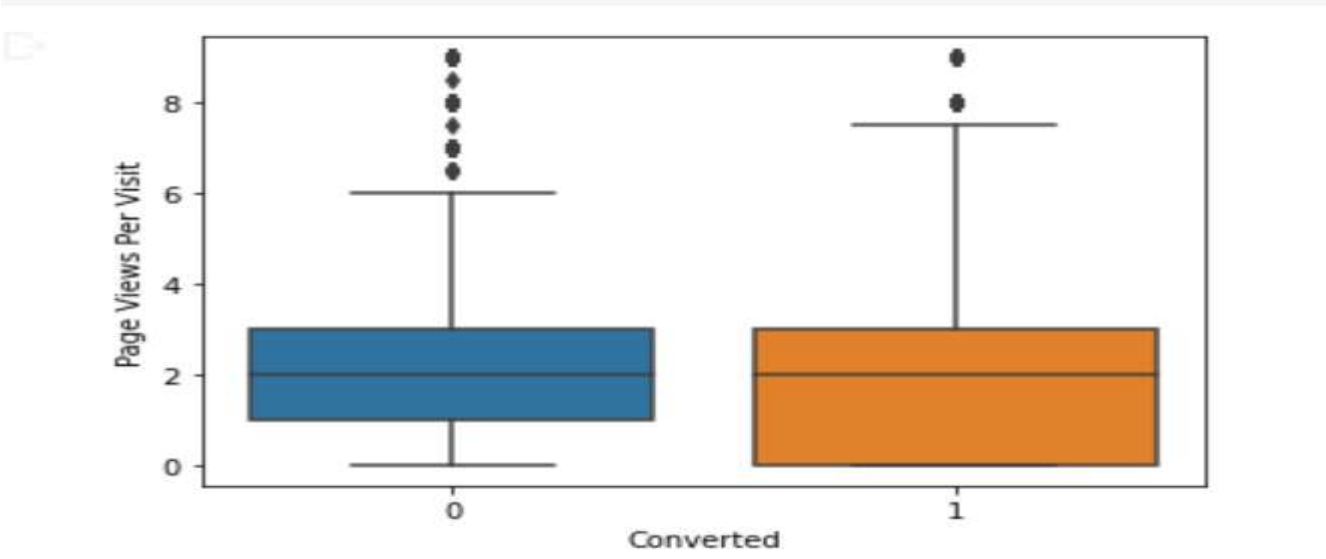
Lead Origin vs Converted



Lead Origin through ‘Lead Add Form’ and ‘Landing Page Submission’ has greater leads conversion rates than others

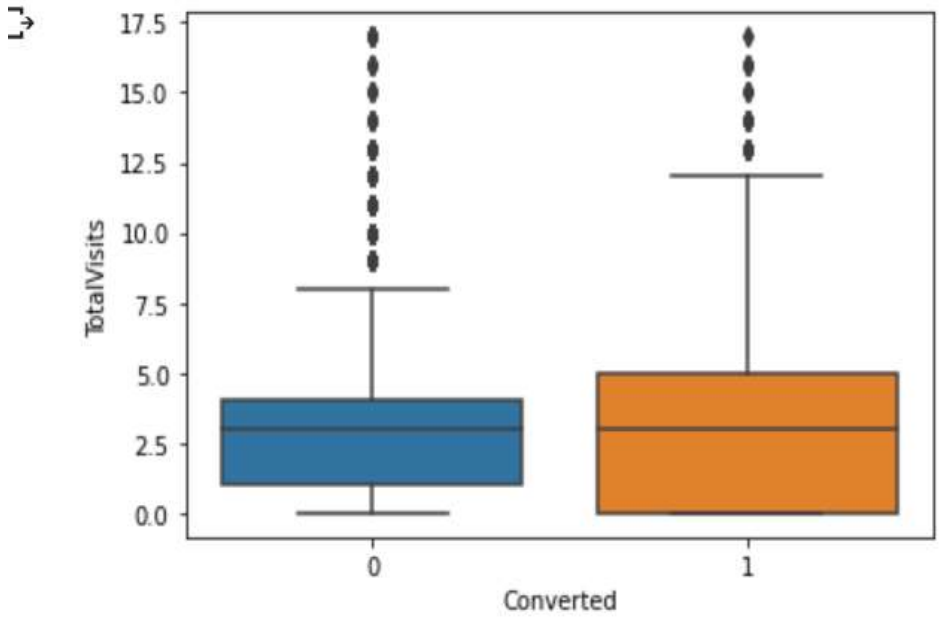
Analyzing Numerical Variables

Page per Visit vs Converted



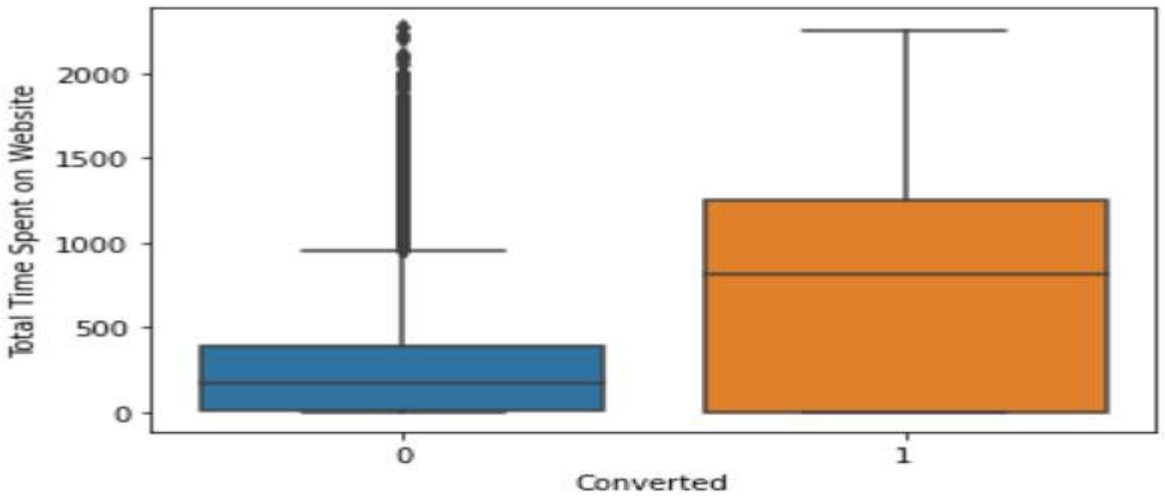
The Upper quartile is longer for the converted customers who has visited the pages maximum times

Total Visits vs Converted



The number of converted customers slightly increases as the total visit count increases

Total Time Spent on Website vs Converted



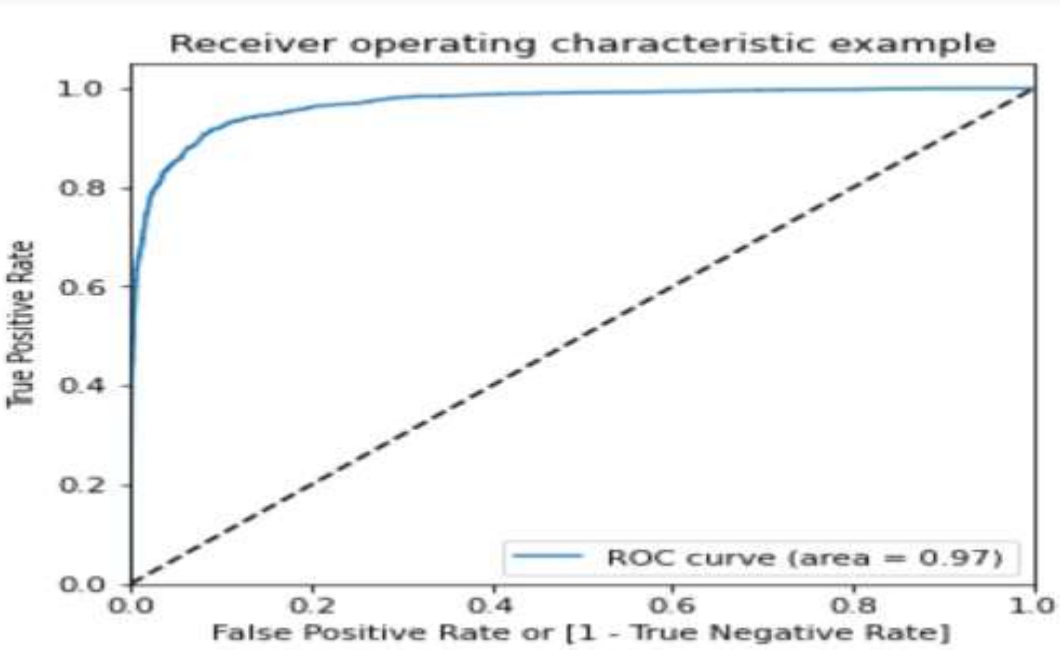
Clearly as total time spent on the website is more so is the conversion rate

Model building Heuristics

TOP-TEN CORRELATION FEATURES OF INDEPENDENT VARIABLES

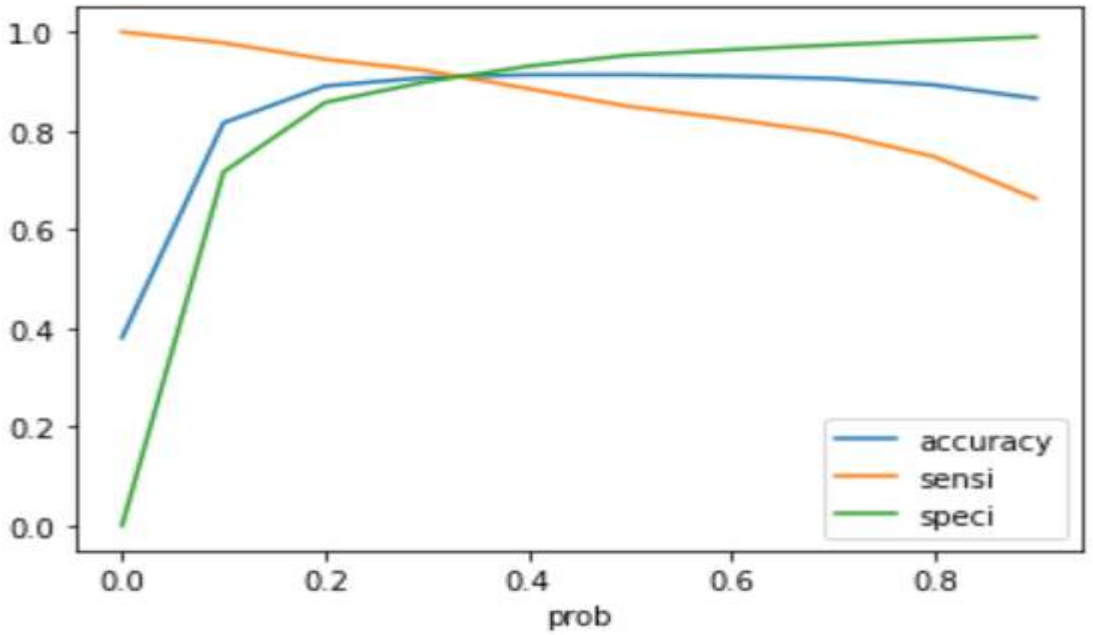
	Column 1	Column 2	Correlation
547	Lead Source_Reference	Lead Add Form	0.845017
182	Page Views Per Visit	TotalVisits	0.721126
1786	Lead Profile_Student of SomeSchool	Tags_Already a student	0.670034
1621	Tags_Will revert after reading the email	Converted	0.654624
675	Last Activity_Email Bounced	Do Not Email	0.626723
410	Lead Source_Direct Traffic	A free copy of Mastering The Interview	0.607037
275	Landing Page Submission	A free copy of Mastering The Interview	0.564394
274	Landing Page Submission	Page Views Per Visit	0.538432
411	Lead Source_Direct Traffic	Landing Page Submission	0.530440
1221	Specialization_management courses	Landing Page Submission	0.505792

ROC Curve for training data



As observed from ROC plot we can justify that our model is performing well as the curve is closer to 0.8

Optimal Cut-off point identification



Here 0.35 is the optimal cut-off point

Confusion metrics parameters of the Training Data

Final Inferences:

Training Dataset: (For the cut-off 0.35----Optimal)

1. Sensitivity : 84%
2. Specificity : 95%
3. Accuracy : 91%

Training Dataset: (For the cut-off 0.5)

1. Sensitivity : 85%
2. Specificity : 95%
3. Accuracy : 91%

We could see that there are no major changes in the values , however as observed before if the business is **stringent about the lead conversion , then we might want to reduce the sensitivity score , so we would have chosen the 0.5 instead of 0.35 as the cut-off 0.5 , the error in specificity measure is less .** However if the business **wants only the accracy measure (i.,e converted should be predicted as converted correctly and vice-versa) then the cut-off 0.35 is precise **

Confusion metric Parameter for Test data

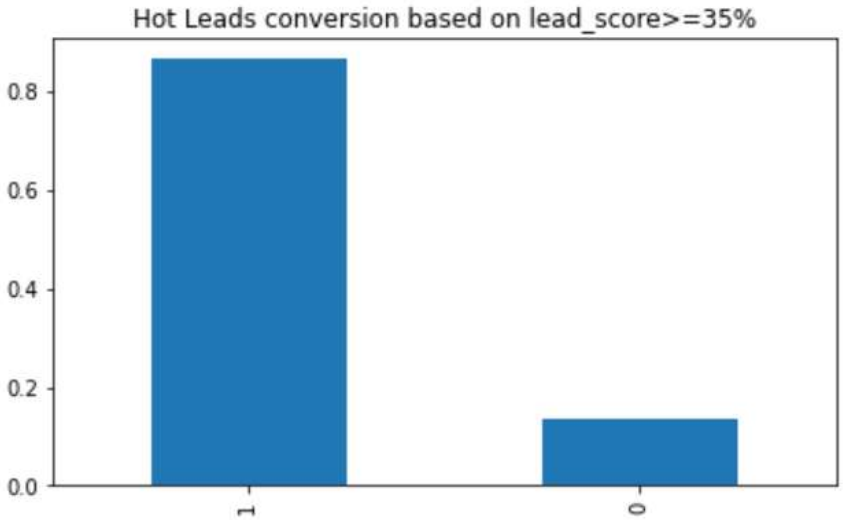
**Testing Dataset: **

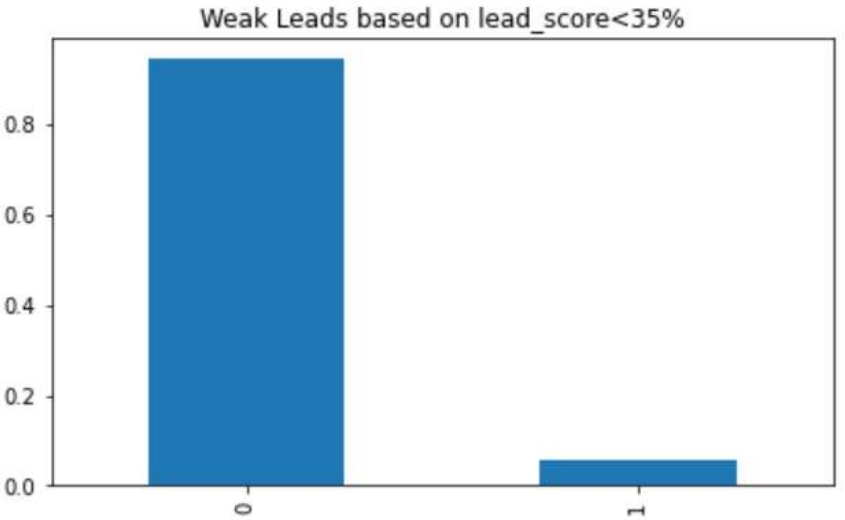
1. Sensitivity : 90%
2. Specificity : 91%
3. Accuracy : 91%

As seen the range of Sensitivity , Specificity and Accuracy remain the same as that of in training data . Our model has been great!

Recommendations to Business

Hot Leads vs Inefficient Customers





Customers with lead score greater than 35 is considered an ideal customer and those with less than 35 are inefficient customers to target

General Recommendations to Business (Dos)

1. We can recommend based on the lead score generated , in the model the hot leads are given the score > 35 and those with less significant given the score <35%

2. Moreover , we can also target the Lead Source channels like ‘Olark Chart’ , ‘Google’

3. We could infer that the management type of courses are having higher lead conversion rates

4. People with occupation as Working-Profession can be targeted and those with currently ‘unemployed’ also show greater conversion rates because they need to industry ready

5. People who visit the website often generally show higher conversion rates (it is also categorised under the lead score)

General Recommendations to Business (Don’t)

1. Any customer whose lead score is less than 30% can be neglected as they are not categorised under hot leads and it is not worth in spending the resources on them.
2. Students generally show less conversion rates might be because of they are hesitant to bear the course fee , so it is good to avoid the students and concentrate with other professional folks.
3. The conversion rate in India is better compared to others , so let us not focus other countries until we see a satisfactory result in India.
4. We can reduce advertising in newspaper, Educational Forums as it does not add any value.
5. We analysed that most of the customers do not like to be called / contacted through mobiles / instead it could be effective if the communication happens through email.
6. The customers don’t show any interest towards supply chain so we can temporarily stop advertising and let us review about the course contents and plan for improvement.

THANK YOU !