Lead Scoring Case Study

Problem Statement

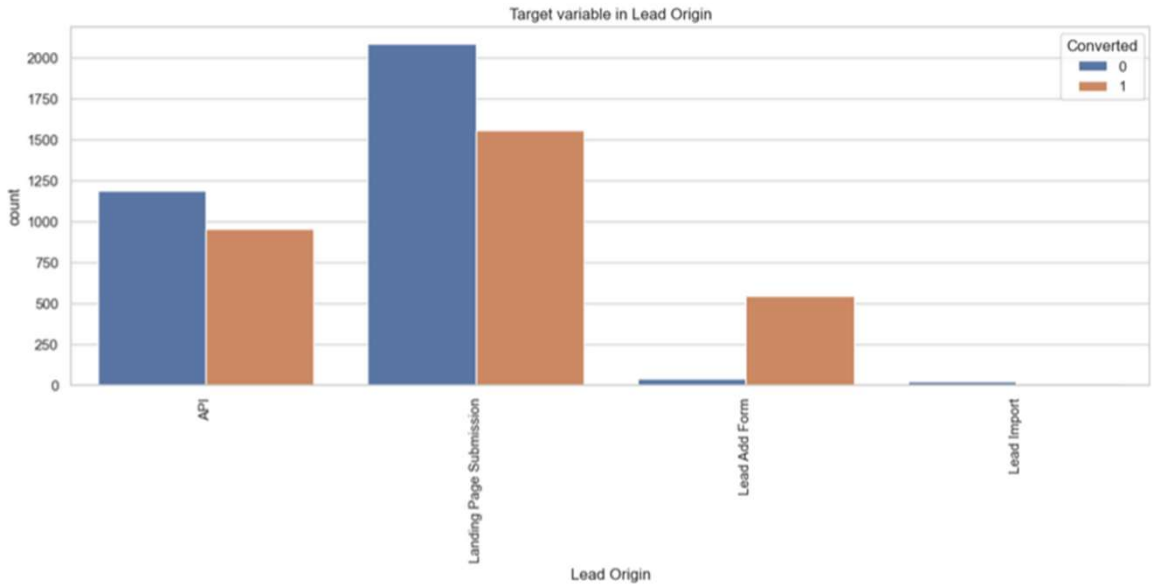
- X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google. Once the leads are acquired, employees from the sales team start making calls, writing emails, etc. so that leads get convert but currently lead conversion rate at X education is around 30%.
- X Education has appointed us to help them to 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 a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.
- The CEO, has given a ballpark of the target lead conversion rate to be around 80%.

Analysis Approach

- Analysis approach for this data set is such a way that,
- Firstly we read the data and then we will clean the data in such a way that we will drop the variable columns with unique data and also drop the columns whose Null value percentage is greater than 45%. And also remove the sales generated variable to avoid ambiguity.
- Then we will check the correlation between the variables and also do the analysis of variables.
- Then we prepare the data by changing the binary variable into “1” or “0”. And also create the dummy variables for the categorical variables.
- Then we will split the data into two parts, (i.e. Test and Train Set). And also do the Re-scaling of the numerical variables.
- Then we build the model with the use of Recursive Features Elimination to select the top 15 important features. And with help of stats model and we try to select the most significant vales and drop the insignificant values.
- Then we find the Optimal Cutoff point with the help of Precision and Recall i.e. 0.4.
- Then we find out that the Conversion percentage of the model based on the Converted variable is aprox 80%.
- Then on basis of model we will implement the learnings to test set and also found that the accuracy will be 78, Sensitivity 79% and Specificity 77%

Insights

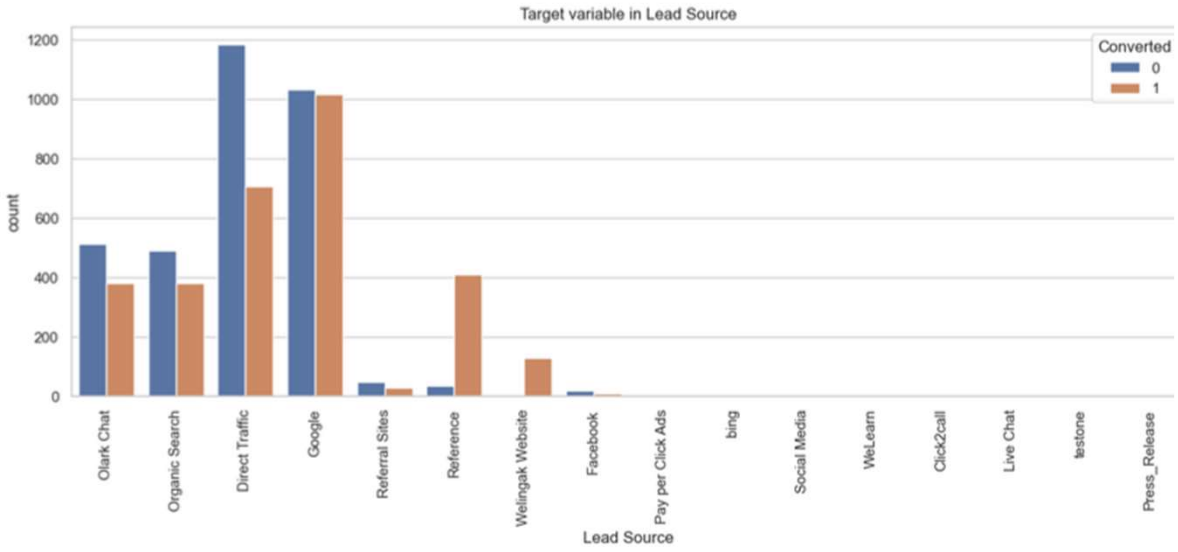
- Lead origin



It has been clearly shows that the Maximum number of leads will be come from the Landing Page Submission but the conversion is more for the API

Insights

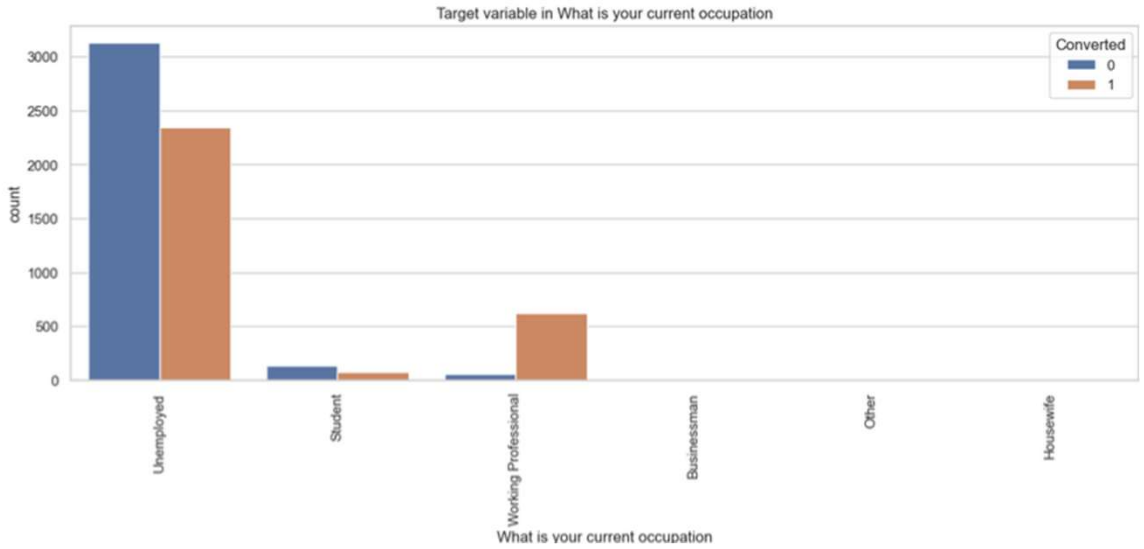
- Lead Source



It has been clearly shows that the Maximum number of leads will be come from the Direct Traffic and Google and there conversion is also hgh.

Insights

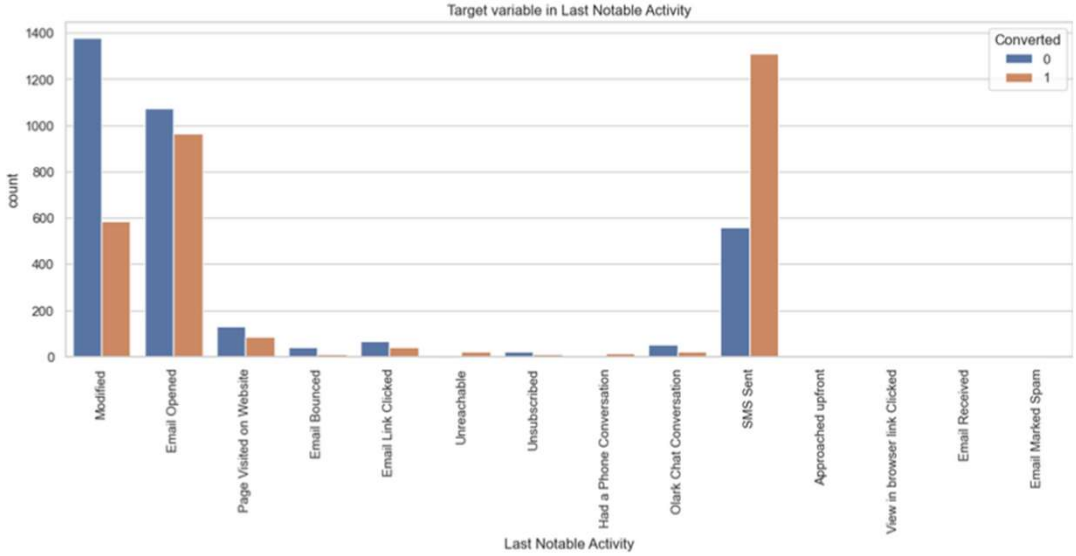
- What is your current Occupation



It has been clearly shows that the Maximum number of leads will be Unemployed but the conversion is very hgh for the working professional..

Insights

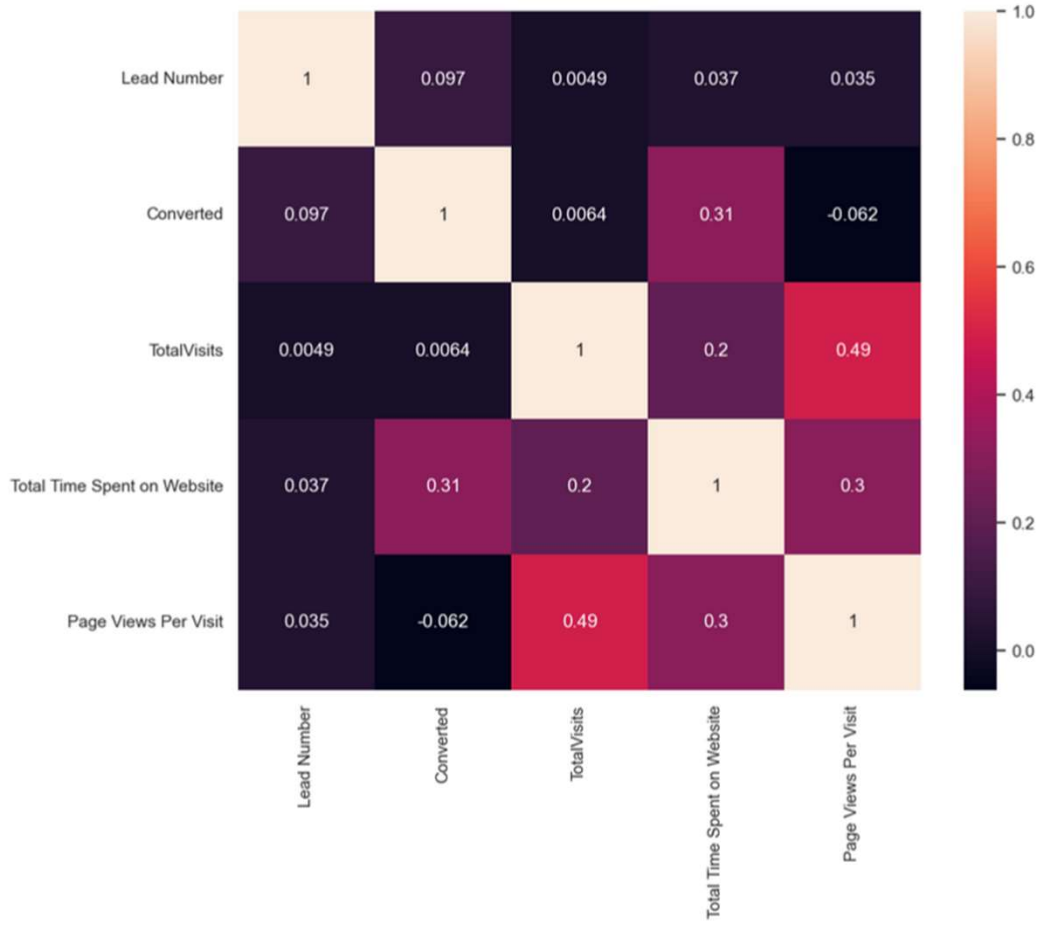
- Last Notable Activity



It has been clearly shows that the Maximum number of leads open the email or SMS but the conversion is very hgh for the SMS Sent.

Insights

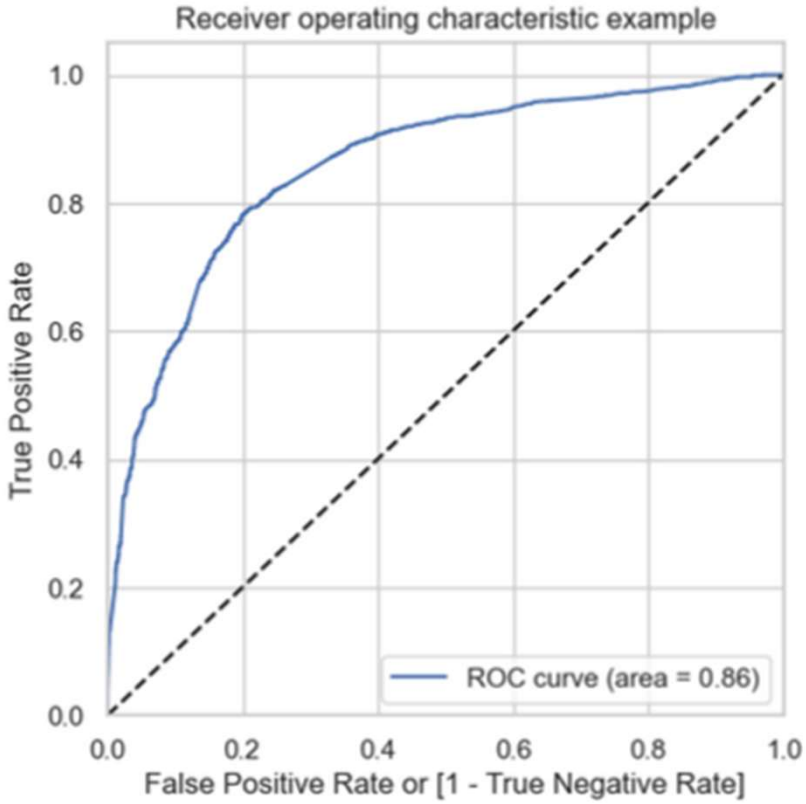
- Correlation



There will be no Correlation between the varia

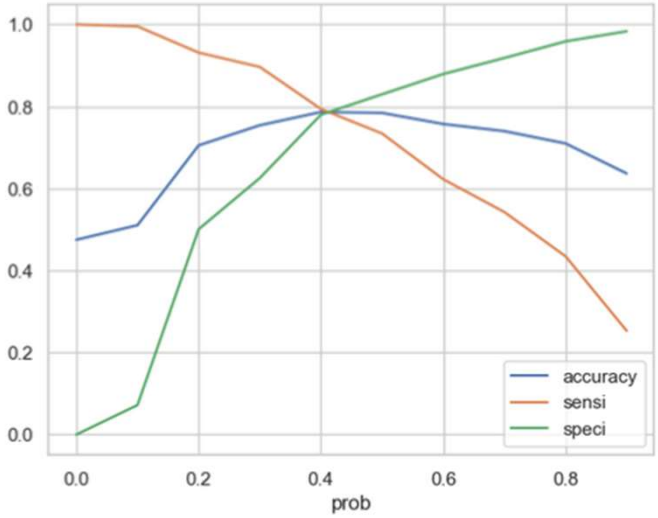
Model Evaluations

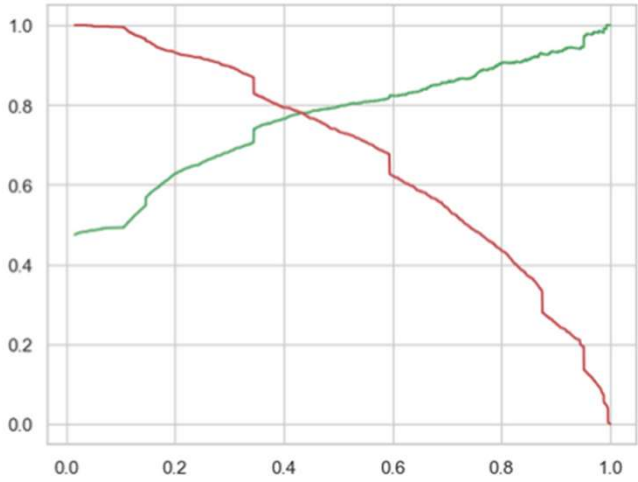
- ROC Curve



The ROC Curve should be close to 1. And we get value of 0.86 it indicates a good predictive model.

Model Evaluations





It has been clearly seen that Optimal Cuttoff value will be 0.4. Thus we can consider any prospect lead with conversion Probabiliy higher than 40% to be hot.

Observation

TRAIN DATA:

Accuracy: 78%

Sensitivity: 79%

Specificity: 77%

TEST DATA:

Accuracy: 79%

Sensitivity: 79%

Specificity: 77%

Final Model Feature list:

a. Lead Source_Olark Chat

b. Lead Origin_Lead Add Form

c. Total Time Spent on Website

d. Last Activity_SMS Sent

e. Last Activity_Olark Chat Conversation

f. What is your current occupation_Working Professionals

g. Last Activity_Email Bounced

h. Last Activity_Had a Phone Conversation

i. Last Notable Activity_Unreachable

Conclusion

A. We see that the conversion rate is 30-35% of the company but after making the model and implementing the learning it shows that the conversion rate has increased drastically to 80%.

B. We see that the Max number of leads are generated by google/direct traffic.

C. Leads who spent more time on website, more likely to convert.

D. We also see that the maximum no. of leads for Last Activity will open the Email and SMS Sent but SMS sent has the high conversion ratio.

E. We also see that the Maximum number of leads are Unemployed but the conversion percentage of working professional is high.