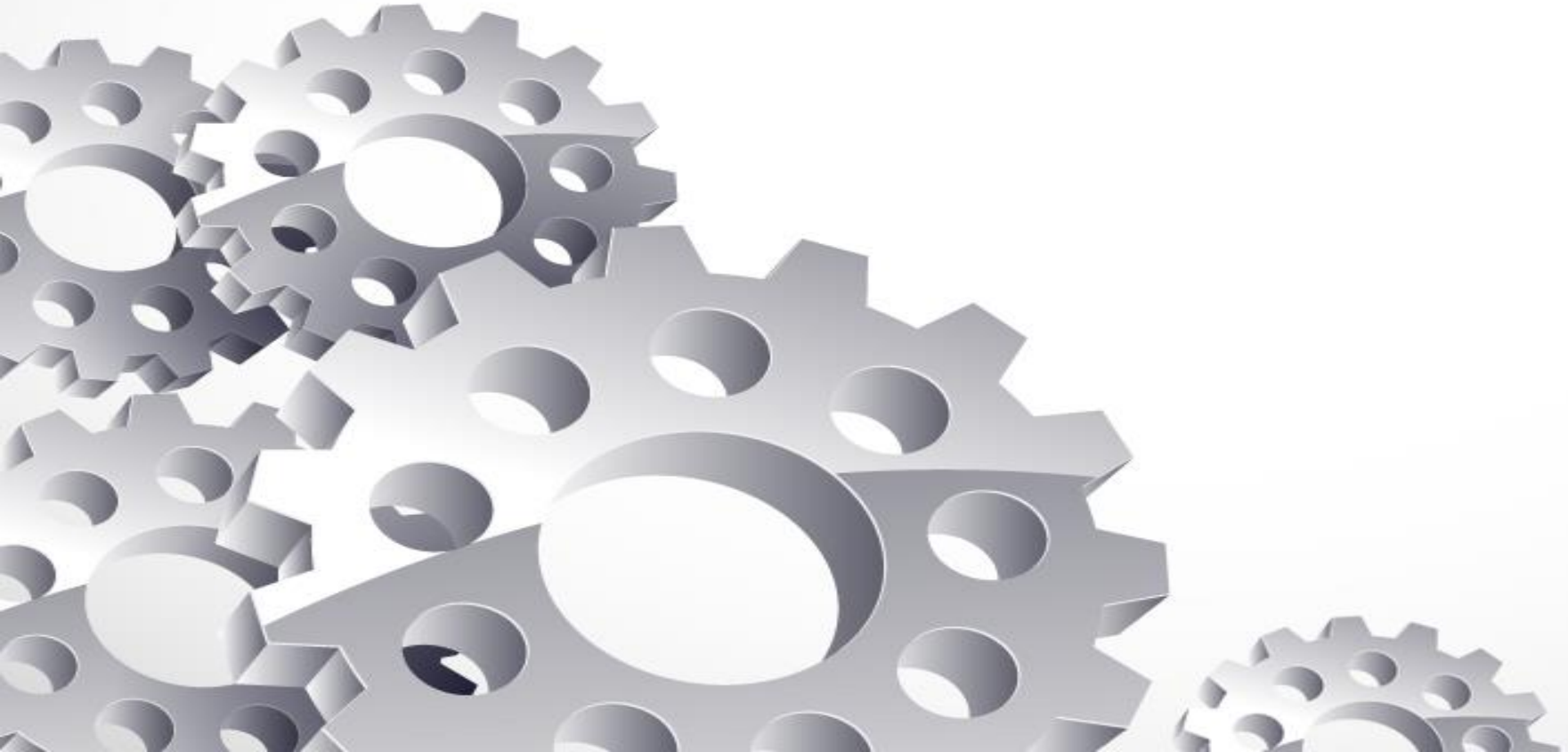
Lead Scoring Case Study



Submitted by:
Raj Ashwani
Anjali M
Shruti Sharma

Problem Statement

- X Education is an Edtech Company which sells online courses to corporate professionals. It gets a lot of leads but the conversion rate is very poor.So,we have to make this process effective, the company wants to identify the most potential leads
- Lead conversion rate will increase if the company is able to get the set of most potential leads
- Sales team will now be more specific & focused on communication part with the promising potential leads

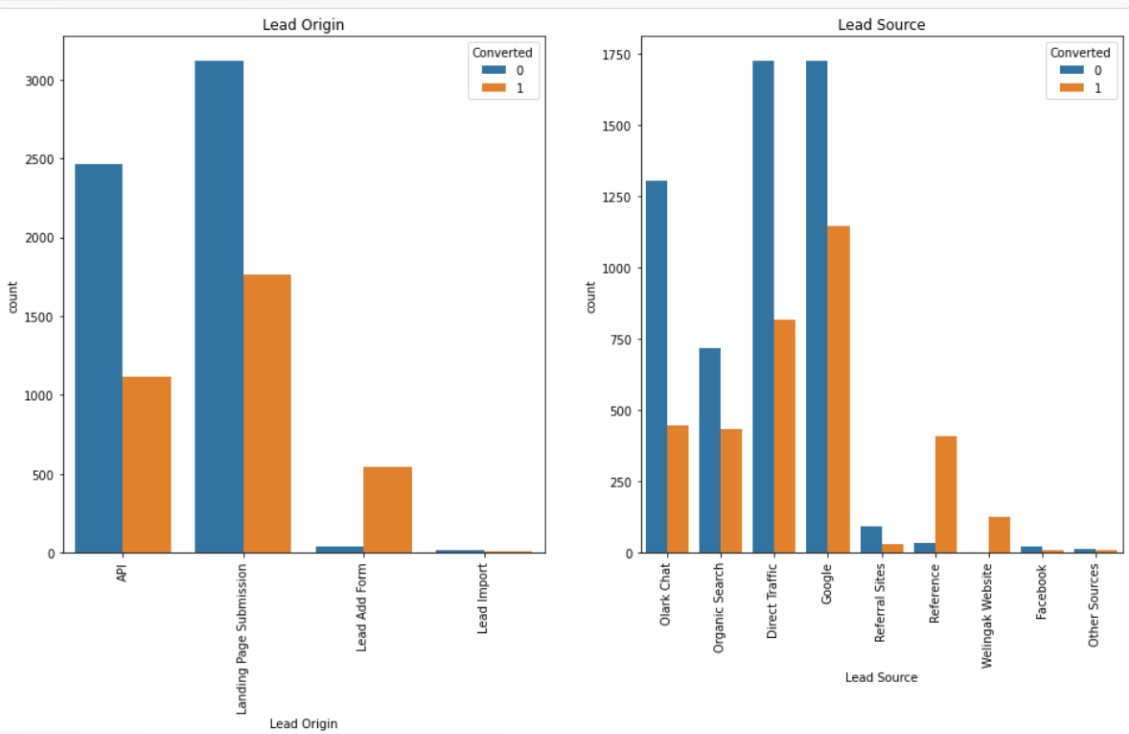
Business Objective

X Education company want us develop a model to help them select the most potential & promising leads. Learners that are more likely to convert into a regular customers. The company requires us to build a model & in that we need to assign the lead score for finding the hot lead .The target lead conversion rate around 80%.

Methodology

- Understanding the data
- Data Preparation
- EDA
- Creation of Dummy Var.
- Train Test Split
- Feature Scaling
- Checked Correlations
- Model Building (Feature Selection Using RFE, improvising the model further inspecting VIF and p- values)
- Built final model
- Model evaluation with different metrics

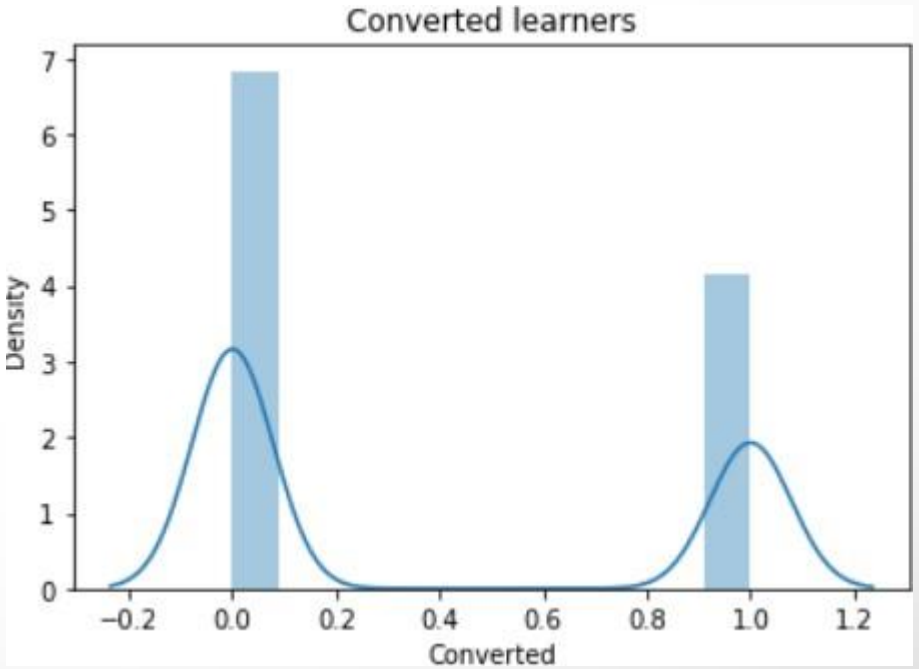
EDA (Univariate Categorical & Continuous Analysis)



Leads & Converted Leads are more on landing pg

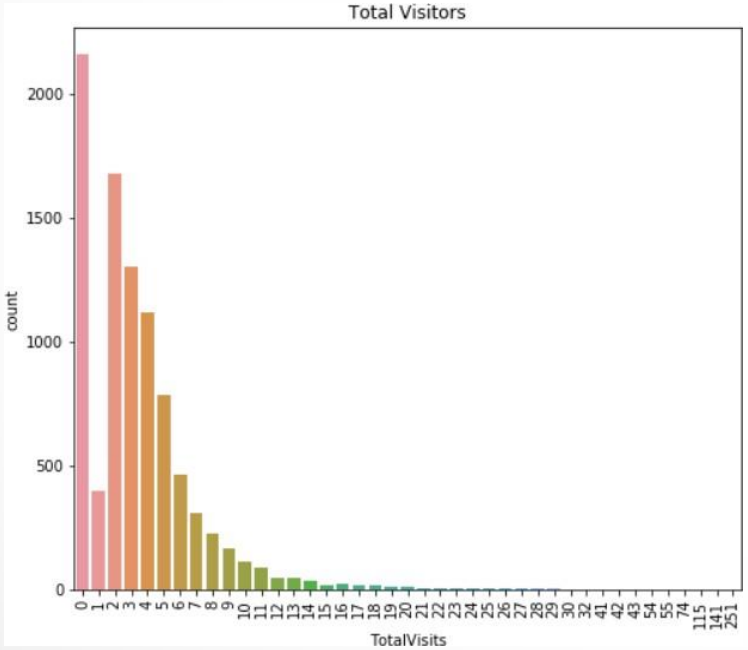
Leads & Converted Leads are more from Google

Comparing not converted & converted
In graph not converted (0) is higher than converted (1)

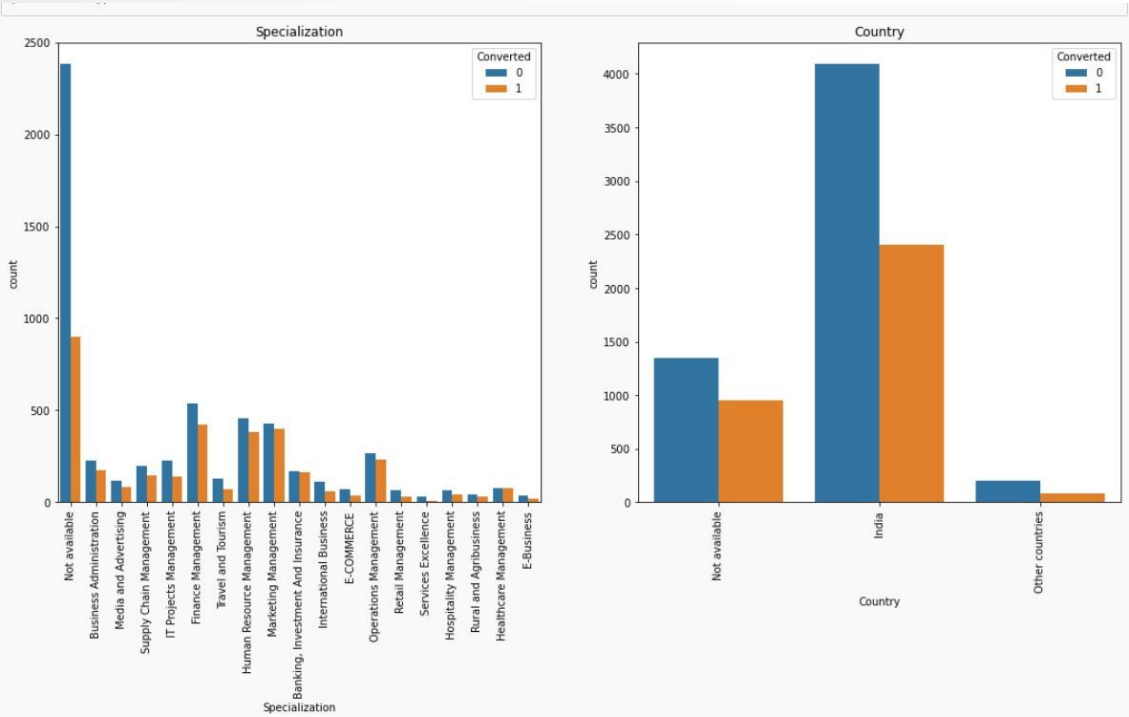


EDA

- Checking and Dropping Null val (more than 37%)
- Given under are website’s most visiting user



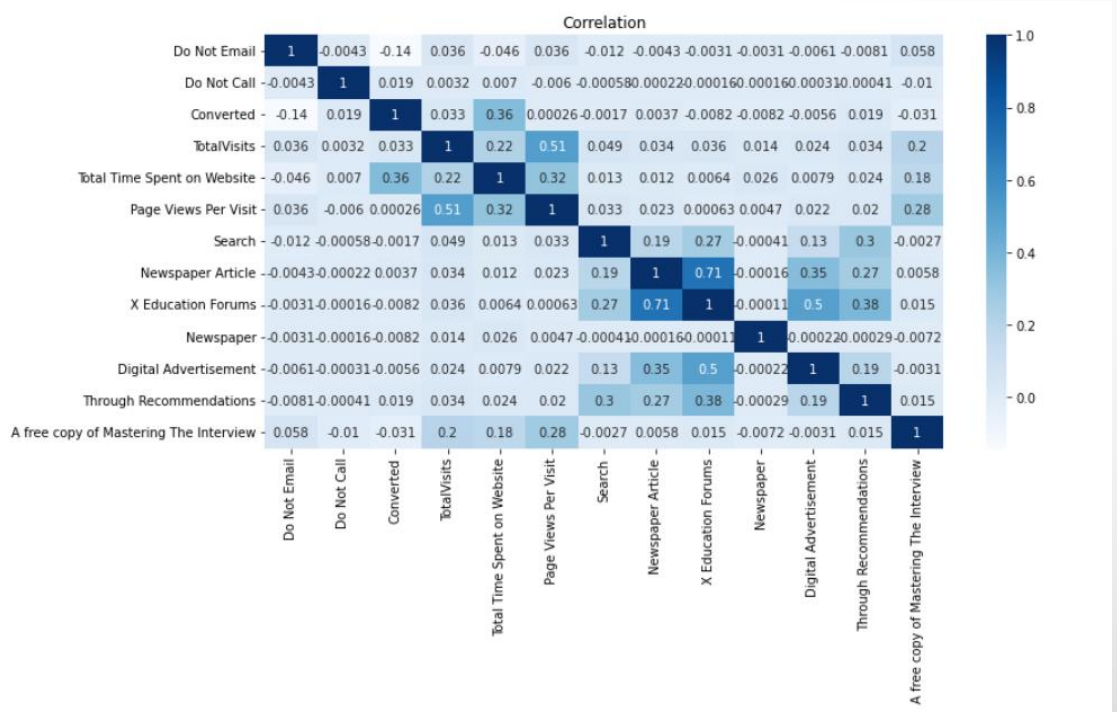
EDA



More Converted Leads- Specialisation 'Not Available'

More Converted Leads- Country India

Correlation- less correlation bet Converted and 'A free copy of mastering the interview'



Model Building

	Converted	Conversion_Prob	Predicted
0	1	0.723125	1
1	0	0.112656	0
2	0	0.638050	1
3	0	0.112656	0
4	0	0.730091	1

Probablity of Conversion & train set prediction

```
# Calculating the sensitivity
TP / float(TP+FN)
```

```
0.6706026058631922
```

```
# calculate the specificity
TN / float(TN+FP)
```

```
0.8842105263157894
```

```
# Precision = TP / TP + FP
TP / (TP + FP)
```

```
0.747895622895623
```

```
#Recall = TP / TP + FN
TP / (TP + FN)
```

```
0.7235342019543974
```

Calculation for model (Precision,Recall etc)

Good Model- Goes till 6 iterations . VIF & P val =<0.05

	Features	VIF
5	Country_Not available	2.57
2	Lead Origin_Landing Page Submission	2.52
7	Specialization_Not available	2.51
0	TotalVisits	2.12
1	Total Time Spent on Website	2.05
3	Lead Origin_Lead Add Form	1.81
8	What is your current occupation_Not available	1.56
4	Lead Source_Welingak Website	1.34
9	What is your current occupation_Working Profes...	1.19
6	Specialization_Hospitality Management	1.02

	Converted	Conversion_Prob	final_predicted
0	0	0.531696	1
1	1	0.653162	1
2	1	0.987691	1
3	1	0.739800	1
4	0	0.135025	0
5	1	0.430266	1
6	0	0.608954	1
7	1	0.554305	1
8	1	0.759387	1
9	0	0.112656	0
10	1	0.206766	0
11	0	0.292566	0
12	0	0.125412	0
13	0	0.133531	0
14	0	0.107415	0
15	1	0.732364	1
16	0	0.170663	0
17	1	0.993245	1
18	1	0.977469	1
19	1	0.843778	1

Taking target variable ‘Converted’ making predictions on test set. Predictions on actual converted & final predicted. According to probability,values in final predicted

.

Conclusion

These variables mattered the most:

- The total time spend on the Company’s website.
- Number of visits.
- When the lead source was Google, Direct traffic, Welingak website etc.
- When the lead origin is Lead add format.
- When their current occupation is as a working professional.

Considering all these points Company could get to know that they have a high oppurtunity to grab almost all of the promising buyers to convert and choose their courses