# LEAD SCORING CASE STUDY FOR X EDUCATION

**AIM:** TO IDENTIFY THE MOST POTENTIAL LEADS, ALSO KNOWN AS 'HOT LEADS'

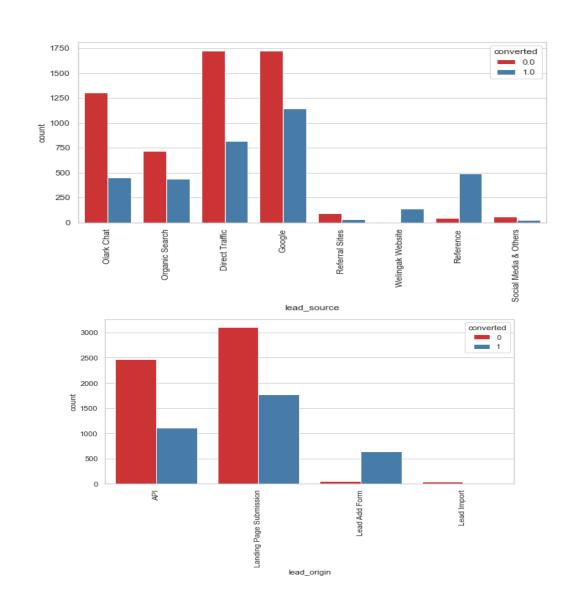
**OBJECTIVE:** WITH ABOUT 37 PREDICTOR VARIABLE PROVIDED WE HAD TO BUILD A LOGISTIC REGRESSION MODEL THAT WILL HELP THE SALES TEAM FOCUS THEIR ENERGY AND TIME TO ONLY THE HIGHLY COVERABLE CANDIDATES. THEREBY INCREASING THE CONVERSION RATE WHICH AT PRESENT IS ONLY 30%.

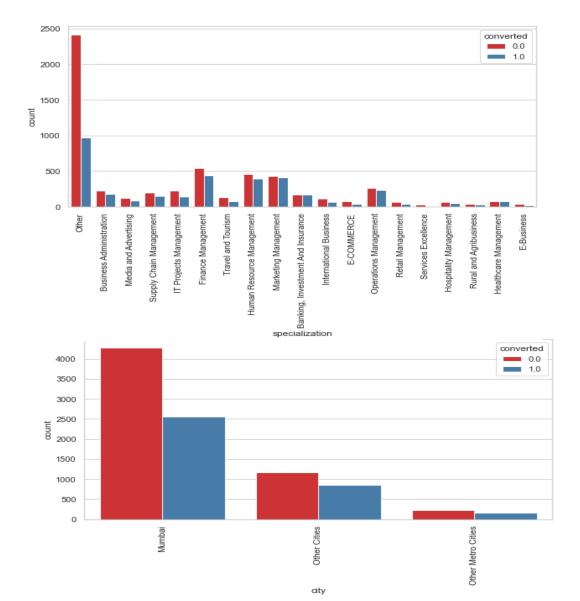
**PROJECT BY:** I. KAMIL SHEIKH

2. SNEHA SOMAN

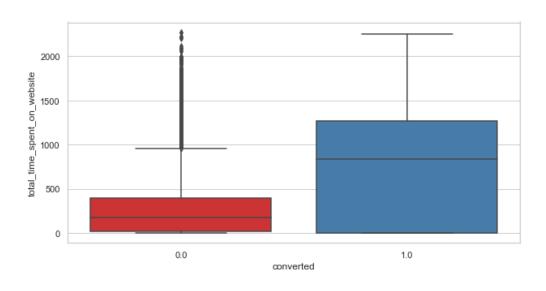
**DATE:** 19<sup>TH</sup> JANUARY 2022

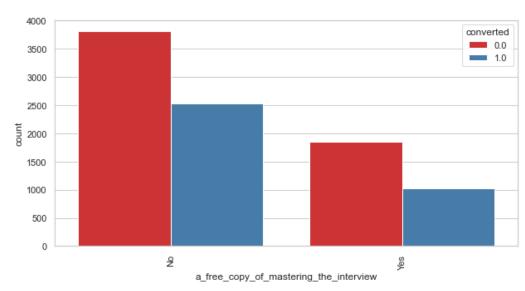
## RETAINED FEATURE VARIABLES AND THEIR RELATION WITH "CONVERTED"

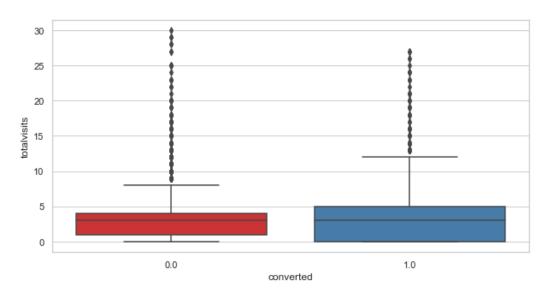


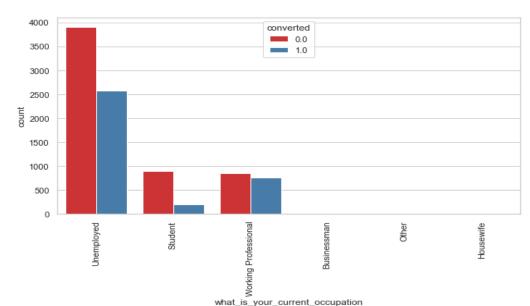


## RETAINED FEATURE VARIABLES AND THEIR RELATION WITH "CONVERTED"









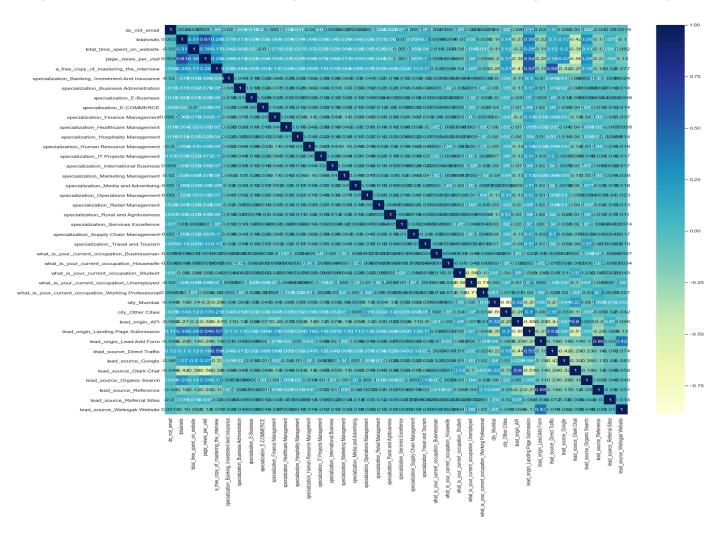
## MAJOR HURDLES DURING DATA CLEANING

- I) The Dataset had many missing values, most of them being categorical in nature.
- 2) There were major data discrepancies in Country and City
- 3) Most of the columns in the data set had very skewed data.
- 4) Some of the key variables used in the prediction model had outliers.

```
df1.country[df1.city=='Mumbai'].value counts()
India
                    2918
gulf country
                      32
United States
other country
                      20
european country
                      11
Name: country, dtype: int64
df1.country[df1.city=='Other Cities of Maharashtra'].value_counts()
India
                    391
gulf country
other country
United States
european country
Name: country, dtype: int64
```

### MACHINE LEANRING MODEL MAKING

After data cleaning, imputing, creating dummy variables and so on we started our model building process with nearly 40 variables. Which was nearly non-readable. There we proceed with RFE directly to ease the process of feature selection.



### MACHINE LEANRING MODEL MAKING

RFE feature selection with 20 cut-off got us 20 most relevant variables with lowest p-value and lowest VIF. VIF was used to study the multi-collinearity of the feature variables.

```
'do_not_email',
                                                    'specialization_Marketing Management',
'total_time_spent_on_website',
                                                    'specialization Media and Advertising',
'specialization_Banking, Investment And Insurance',
                                                    'specialization_Operations Management',
'specialization_Business Administration',
                                                    'specialization Rural and Agribusiness',
'specialization_E-COMMERCE',
                                                    'specialization_Services Excellence',
'specialization_Finance Management',
                                                    'specialization Supply Chain Management',
                                                    'what_is_your_current_occupation_Housewife',
'specialization_Healthcare Management',
'specialization Human Resource Management',
                                                    'what_is_your_current_occupation_Student',
'specialization_IT Projects Management',
                                                    'lead_origin_Landing Page Submission',
'lead_origin_Lead Add Form',
                                                    'lead_source_Welingak Website'
```

## **MACHINE LEANRING MODEL MAKING**

## Heatmap reflecting the co-relation between each of the variables.

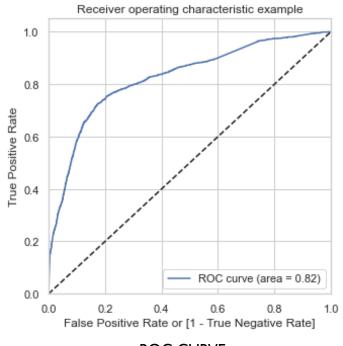
do_not_email	1	-0.037	-0.022	0.011	0.023	0.00043	0.014	-0.02	0.07	-0.027	-0.0036	-0.0063	0.027	0.075	-0.0051	0.034	0.11	-0.036	0.0019
total_time_spent_on_website	-0.037	1	0.042	0.062	-0.01	0.078	0.02	0.07	0.025	0.068	0.038	0.05	0.025	0.0036	0.052	-0.084	0.28	-0.14	-0.082
specialization_Banking, Investment And Insurance	-0.022	0.042	1	-0.043	-0.022	-0.066	-0.025	-0.062	-0.039	-0.062	-0.03	-0.047	-0.018	-0.012	-0.038	-0.0044	0.1	0.01	-0.024
specialization_Business Administration	0.011	0.062	-0.043	1	-0.024	-0.075	-0.029	-0.069	-0.044	-0.07	-0.033	-0.052	-0.021	-0.013	-0.043	-0.032	0.13	-0.012	-0.014
specialization_E-COMMERCE	0.023	-0.01	-0.022	-0.024	1	-0.038	-0.015	-0.035	-0.022	-0.036	-0.017	-0.027	-0.01	-0.0068	-0.022	-0.003	0.084	-0.021	-0.014
specialization_Finance Management	0.00043	0.078	-0.066	-0.075	-0.038	1	-0.045	-0.11	-0.068	-0.11	-0.052	-0.082	-0.032	-0.021	-0.067	-0.035	0.19	0.016	-0.037
specialization_Healthcare Management	0.014	0.02	-0.025	-0.029	-0.015	-0.045	1	-0.042	-0.026	-0.042	-0.02	-0.031	-0.012	-0.008	-0.026	-0.0087	0.057	0.022	-0.0058
specialization_Human Resource Management	-0.02	0.07	-0.062	-0.069	-0.035	-0.11	-0.042	1	-0.063	-0.1	-0.048	-0.076	-0.03	-0.019	-0.062	-0.025	0.16	0.038	-0.016
specialization_IT Projects Management	0.07	0.025	-0.039	-0.044	-0.022	-0.068	-0.026	-0.063	1	-0.064	-0.031	-0.048	-0.019	-0.012	-0.039	-0.016	0.14	-0.036	-0.017
specialization_Marketing Management	-0.027	0.068	-0.062	-0.07	-0.036	-0.11	-0.042	-0.1	-0.064	1	-0.049	-0.077	-0.03	-0.02	-0.063	-0.042	0.13	0.029	-0.039
specialization_Media and Advertising	-0.0036	0.038	-0.03	-0.033	-0.017	-0.052	-0.02	-0.048	-0.031	-0.049	1	-0.036	-0.014	-0.0093	-0.03	-0.025	0.11	-0.036	-0.018
specialization_Operations Management	-0.0063	0.05	-0.047	-0.052	-0.027	-0.082	-0.031	-0.076	-0.048	-0.077	-0.036	1	-0.022	-0.015	-0.047	-0.023	0.12	-0.014	-0.023
specialization_Rural and Agribusiness	0.027	0.025	-0.018	-0.021	-0.01	-0.032	-0.012	-0.03	-0.019	-0.03	-0.014	-0.022	1	-0.0057	-0.018	0.0057	0.064	-0.021	-0.011
specialization_Services Excellence	0.075	0.0036	-0.012	-0.013	-0.0068	-0.021	-0.008	-0.019	-0.012	-0.02	-0.0093	-0.015	-0.0057	1	-0.012	0.016	0.042	-0.0079	-0.0074
specialization_Supply Chain Management	-0.0051	0.052	-0.038	-0.043	-0.022	-0.067	-0.026	-0.062	-0.039	-0.063	-0.03	-0.047	-0.018	-0.012	1	-0.024	0.13	-0.022	-0.017
what_is_your_current_occupation_Student	0.034	-0.084	-0.0044	-0.032	-0.003	-0.035	-0.0087	-0.025	-0.016	-0.042	-0.025	-0.023	0.0057	0.016	-0.024	1	-0.051	-0.082	-0.045
lead_origin_Landing Page Submission	0.11	0.28	0.1	0.13	0.084	0.19	0.057	0.16	0.14	0.13	0.11	0.12	0.064	0.042	0.13	-0.051	1	-0.31	-0.13
lead_origin_Lead Add Form	-0.036	-0.14	0.01	-0.012	-0.021	0.016	0.022	0.038	-0.036	0.029	-0.036	-0.014	-0.021	-0.0079	-0.022	-0.082	-0.31	1	0.42
lead_source_Welingak Website	0.0019	-0.082	-0.024	-0.014	-0.014	-0.037	-0.0058	-0.016	-0.017	-0.039	-0.018	-0.023	-0.011	-0.0074	-0.017	-0.045	-0.13	0.42	1
	do_not_email	total_time_spent_on_website	ecialization_Banking, Investment And Insurance	specialization_Business Administration	specialization_E-COMMERCE	specialization_Finance Management	specialization_Healthcare Management	specialization_Human Resource Management	specialization_IT Projects Management	specialization_Marketing Management	specialization_Media and Advertising	specialization_Operations Management	specialization_Rural and Agribusiness	specialization_Services Excellence	specialization_Supply Chain Management	what_is_your_current_occupation_Student	lead_origin_Landing Page Submission	lead_origin_Lead Add Form	lead_source_Welingak Websile

### FINAL MACHINE LEARNING MODEL FOR X education SALES TEAM

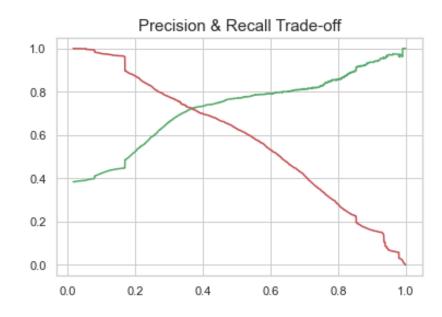
With a probability threshold of 0.30 we were able to achieve the following score which when tested on the test data set matched well with the train data set. Indicating that it is a dependable model.

TRAI	N DATA	TEST DATA								
Confusion matrics:	2982     983       554     1916	Confusion matrics:	1261     437       233     828							
Accuracy score:	<b>76</b> %	Accuracy score:	75.71%							
Sensitivity score:	77.57%	Sensitivity score:	78.03%							
Specificity score:	75.20%	Specificity score:	74.26%							
False Positive Rate:	24.79%									

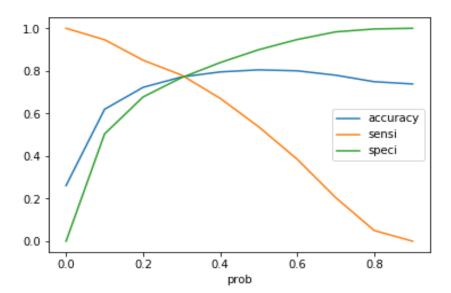
## FINAL MACHINE LEARNING MODEL FOR X education SALES TEAM



**ROC CURVE** 







### **SUGGESTION**

#### **GENERAL SUGGESTION**

- I. The initial data collection form needs some major rework
- 2. More options on providing various free reading or self preparing materials should be provided to gain the customer attention and interest.
- 3. Even though customers select 'NO' for call & e-mail they are highly convertible.

### **SPECIALIZATION BASED SUGGESTION**

- I. Banking, Investment And Insurance 1.1235
- 2. Business Administration 0.8662
- 3. E-Commerce 0.8385
- 4. Finance Management 0.8982
- 5. Healthcare Management 1.1608
- 6. Human Resource Management 0.9038
- 7. IT Projects Management 0.8173
- 8. Marketing Management 0.9700
- 9. Media and Advertising 0.8955
- 10. Operations Management 1.0477
- 11. Rural and Agribusiness 0.8370
- 12. Services Excellence 1.2927
- 13. Supply Chain Management 0.9176