# Lead Scoring Case Study

Priyal Verma

## AGENDA

**Problem Statement** 

Primary goals

Strategy

Methodology

**EDA** 

**Correlation Matrix** 

Outlier

**Modal Evaluation** 

Conclusion

PRESENTATION TITLE

## Problem Statement

X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google, Microsoft.

Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their details like email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals.

Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not.

## PRIMARY GOALS

- The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance
- > The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

## **STRATEGY**

- Source the data for analysis
- Clean and treat missing values
- Exploratory Data Analysis on the data .
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics -Specificity and Sensitivity or Precision and Recall.
- Applying the best model in Test data

SENTATION TITLE

## METHODOLOGY

Data Sourcing, Cleaning and Preparation



Feature Scaling and Splitting Train and Test Sets

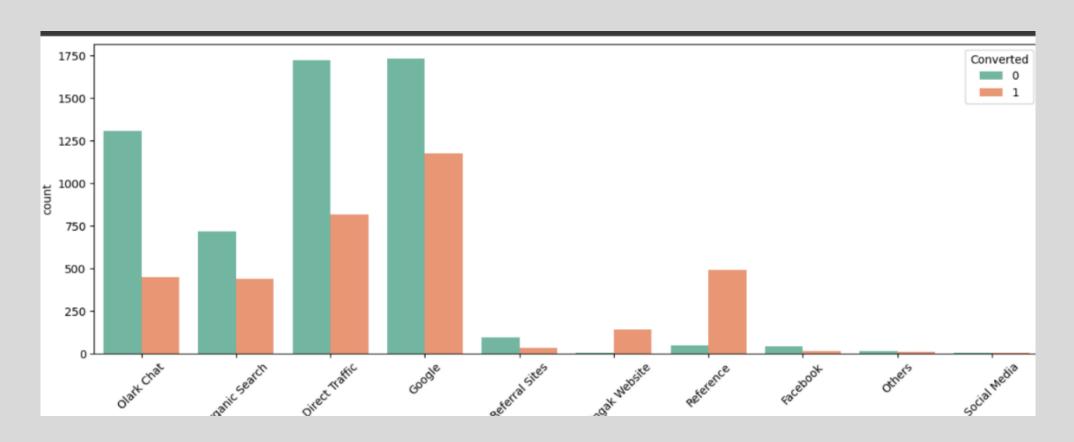


**Model Building** 



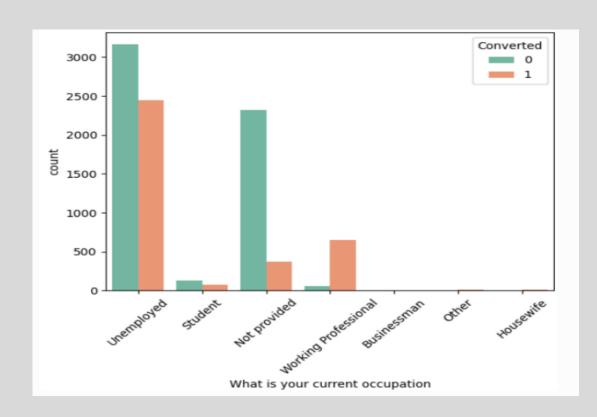
Result

### **Exploratory Data Analysis**

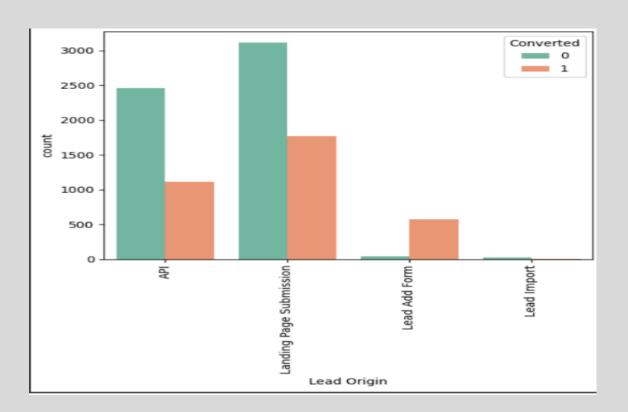


- Maximum Leads are generated by Google and Direct Traffic
- Conversion rate of Reference leads and Welinkgak Website leads is very high.

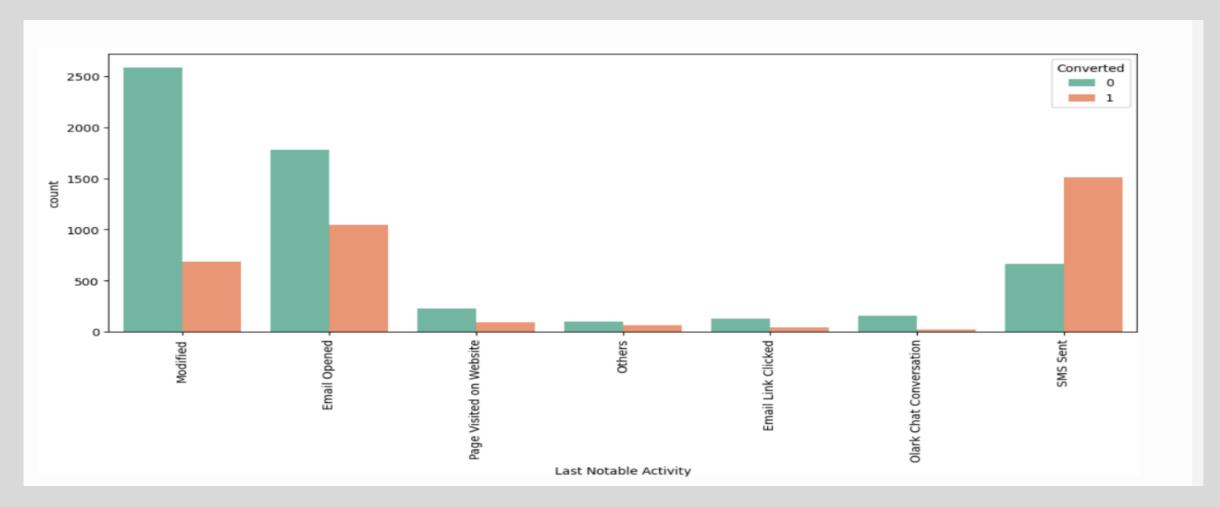
#### **ANALYSIS**



Maximum leads generated are unemployed and their conversion rate is more than 50%. Conversion rate of working professionals is very high.

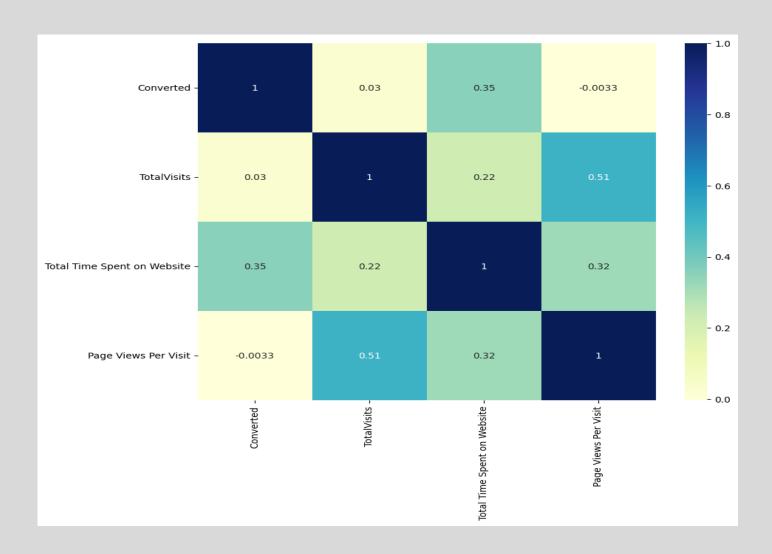


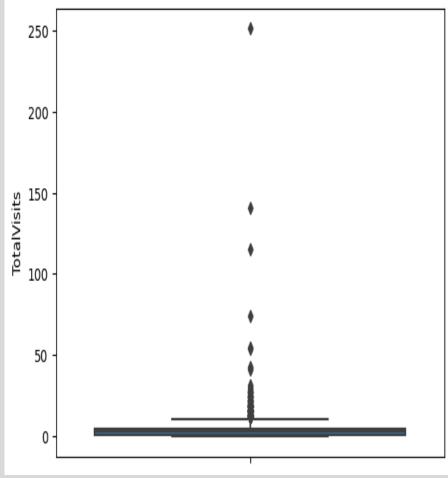
Maximum leads generated are from Landing Page Submission



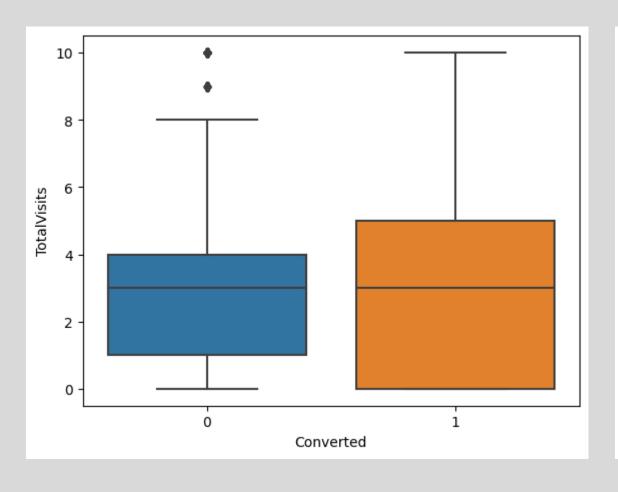
- > Maximum leads are generated having last activity as Email opened but conversion rate is not too good.
- > SMS sent as last activity has high conversion rate.

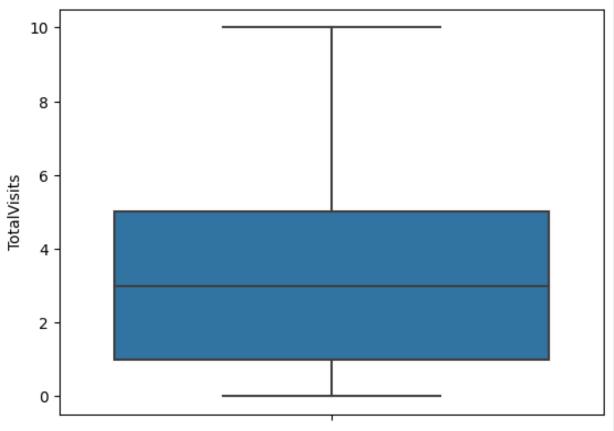
#### Checking correlations of numeric values using heatmap





#### **Outlier Treatment**

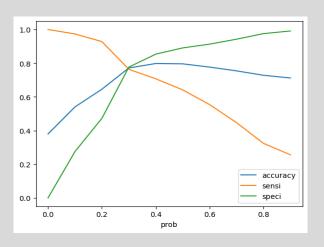


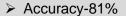


#### **CORRELATION MATRIX**

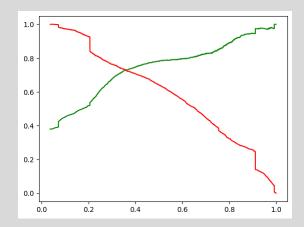
	Do Not Email -	- 1	-0.14	0.00083	3-0.048	0.02	0.096	-0.022	-0.017	0.11	-0.017	-0.067	-0.052	0.034	-0.036	-0.01	-0.0044	0.0025	-0.0087	-0.0092	-0.0019	0.0069	-0.022	-0.048
	Converted -	-0.14	1	0.047		-0.0038	-0.041	0.31	-0.011	-0.075	-0.011	0.034	-0.13	-0.0019	0.25	-0.032	0.0037	0.15	0.015	0.04	0.018 -	0.0056	0.13	0.32
	TotalVisits -	0.00083	3 0.047	1	0.36	0.77	0.46	-0.28	-0.061	0.13	-0.059	0.21	-0.5	0.3	-0.24	0.091	-0.0036	-0.13	0.015	-0.0083	0.034	-0.032	0.077 -	0.00098
1.	Total Time Spent on Website -	-0.048		0.36	1	0.37	0.29	-0.19	-0.051	0.14	-0.05	0.21	-0.38	0.1	-0.16	0.013	0.0019	-0.092	0.013	0.007	0.032	-0.013	0.084	0.097
1	Page Views Per Visit -	0.02	-0.0038	0.77	0.37	1		-0.32	-0.067	0.16	-0.063	0.26	-0.58	0.31	-0.27	0.079	0.002	-0.14	0.0098	-0.0075	0.02	-0.037	0.096	-0.0087
/	Lead Origin_Landing Page Submission -	0.096	-0.041	0.46	0.29	0.56	1	-0.29	-0.063	0.52	-0.059	0.074	-0.53	0.052	-0.24	-0.076	-0.0011	-0.13	-0.017	0.0012	0.0052-0	0.00016	0.091	-0.0021
	Lead Origin_Lead Add Form -	-0.022	0.31	-0.28	-0.19	-0.32	-0.29	1	-0.016	-0.17	-0.016	-0.16	-0.13	-0.1	0.85	-0.032	-0.004		0.0069	0.034	-0.011	0.0094	0.056	0.19
	Lead Origin_Lead Import -	-0.017	-0.011	-0.061	-0.051	-0.067	-0.063	-0.016	1	-0.036	0.97	-0.036	-0.029	-0.022	-0.013	-0.0069	-0.00087	-0.007	-0.0017	-0.0018	-0.0024	-0.0089	0.028	-0.0022
	Lead Source_Direct Traffic -	0.11	-0.075	0.13	0.14	0.16		-0.17	-0.036	1	-0.036	-0.43	-0.3	-0.24	-0.14	-0.073	-0.0092	-0.075	-0.01	-0.012	0.017	0.031	0.044	-0.009
	Lead Source_Facebook -	-0.017	-0.011	-0.059	-0.05	-0.063	-0.059	-0.016	0.97	-0.036	1	-0.04	-0.029	-0.022	-0.013	-0.0069	-0.00087	-0.007	-0.0017	-0.0018-	-0.0024	0.0037	0.024	-0.0022
	Lead Source_Google -	-0.067	0.034	0.21	0.21	0.26	0.074	-0.16	-0.036	-0.43	-0.04	1	-0.33	-0.26	-0.15	-0.081	-0.01	-0.082	-0.00440	0.00098-	0.0045	-0.033	0.033	-0.015
	Lead Source_Olark Chat -	-0.052	-0.13	-0.5	-0.38	-0.58	-0.53	-0.13	-0.029	-0.3	-0.029	-0.33	1	-0.19	-0.11	-0.058	-0.0072	-0.059	-0.0051	-0.015	-0.013	0.023	-0.16	-0.085
	Lead Source_Organic Search -	0.034	-0.0019		0.1	0.31	0.052	-0.1	-0.022	-0.24	-0.022	-0.26	-0.19	1	-0.086	-0.045	-0.0056	-0.046	0.022	0.009	0.0089	-0.023	0.044	-0.0039
	Lead Source_Reference -	-0.036	0.25	-0.24	-0.16	-0.27	-0.24	0.85	-0.013	-0.14	-0.013	-0.15	-0.11	-0.086	1	-0.027	-0.0034	-0.027	0.011	0.042	0.0092	0.017	0.0013	0.24
	Lead Source_Referral Sites -	-0.01	-0.032	0.091	0.013	0.079	-0.076	-0.032	-0.0069	-0.073	-0.0069	-0.081	-0.058	-0.045	-0.027	1	-0.0017	-0.014	-0.0035	-0.0037	0.0048	-0.018	-0.017	-0.0048
	Lead Source_Social Media -	-0.0044	0.0037	-0.0036	0.0019	0.002 -	0.0011	-0.004	0.00087	7-0.0092·	0.00087	-0.01	-0.0072	-0.0056	-0.0034	-0.0017	1	-0.0018	0.00044	0.00047	-0.0006	0.0023	0.012	-0.0042
	Lead Source_Welingak Website -	0.0025	0.15	-0.13	-0.092	-0.14	-0.13	0.45	-0.007	-0.075	-0.007	-0.082	-0.059	-0.046	-0.027	-0.014	-0.0018	1	-0.0036	-0.0038	0.0049	-0.012	0.095	-0.034
	What is your current occupation_Businessman -	-0.0087	0.015	0.015	0.013	0.0098	-0.017	0.0069	-0.0017	-0.01	-0.0017	-0.0044	-0.0051	0.022	0.011	-0.0035	-0.00044	-0.0036	1	0.00093	-0.0012	0.0045	-0.037	-0.0084
	What is your current occupation_Housewife -	-0.0092	0.04	-0.0083	0.007	-0.0075	0.0012	0.034	0.0018	-0.012	-0.0018	0.00098	-0.015	0.009	0.042	-0.0037	-0.00047	-0.0038	0.00093	1	-0.0013	0.0048	-0.039	-0.0089
	What is your current occupation_Other -	-0.0019	0.018	0.034	0.032	0.02	0.0052	-0.011	0.0024	0.017	-0.0024	-0.0045	-0.013	0.0089	-0.0092	-0.0048	-0.0006	-0.0049	-0.0012	-0.0013	1	0.0062	-0.05	-0.012
	What is your current occupation_Student -	0.0069	-0.0056	-0.032	-0.013	-0.037-0	0.0001€	50.0094	-0.0089	0.031	0.0037	-0.033	0.023	-0.023	0.017	-0.018	-0.0023	-0.012	-0.0045	-0.0048	-0.0062	1	-0.19	-0.043
	What is your current occupation_Unemployed -	-0.022	0.13	0.077	0.084	0.096	0.091	0.056	0.028	0.044	0.024	0.033	-0.16	0.044	0.0013	-0.017	0.012	0.095	-0.037	-0.039	-0.05	-0.19	1	-0.35
	What is your current occupation_Working Professional -				0.097	-0.0087	0.0021	0.19	0.0022	-0.009	-0.0022	-0.015		-0.0039	0.24	-0.0048	-0.0042		-0.0084	-0.0089	-0.012	-0.043		1
		Do Not Email	Converted	TotalVisits	on Website	Per Visit	Lead Origin_Landing Page Submission	Add Form	Lead Origin_Lead Import	ct Traffic	Lead Source_Facebook	Lead Source_Google	ark Chat	Lead Source_Organic Search	eference	rral Sites	ial Media	Website	nessman	ousewife	on_Other	_Student	What is your current occupation_Unemployed	What is your current occupation_Working Professional
		Do	8	F	pent on	Page Views I	<sup>a</sup> ge Su	Lead Origin_Lead Add	gin_Lea	Source_Direct	ource	Source	Lead Source_Olark	_Organi	Lead Source_Refer	Lead Source_Referral	Lead Source_Social	Lead Source_Welingak	on_Busi	ation_Ho	occupation	upation	on_une	cing Pro
					Time Spent	Pag	anding F	d Origir	ead Ori	Lead Sour	Lead S	Leac	Lead So	Source	Lead S	ad Sour	ad Soul	ource_M	ccupati	t occupatior	irrent o	current occupatior	ccupati	on_Work
					Total		origin_La	Lea	_	의				Lead		Lei	3	Lead So	urrent o	r current	your cu	our curr	urrent o	ccupati
							Lead 0												What is your current occupation_Busi	What is your	What is your	What is your o	s your c	irrent o
																			What is	Wha		≫	What is	your cu
																								What is

## Model Evaluation -Sensitivity and Specificity on Train Data Set





- > Sensitivity-80%
- > Specificity-82%
- ➤ FalsePositiveRate-18%
- ➤ PositivePredictiveValue-74%
- ➤ PositivePredictiveValue—86%



> Precision-79%

➤ Recall-71%

#### CONCLUSION

- 1 -While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.
- 2 -Accuracy, Sensitivity and Specificity values of test set are around 81%, 79% and 82% which are approximately closer to the respective values calculated using trained set.
- 3- Also the lead score calculated shows the conversion rate on the final predicted model is around 80% (in train set) and 79% in test set
- 4 The top 3 variables that contribute for lead getting converted in the model are-

Total time spent on website

Lead Add Form from Lead Origin

Had a Phone Conversation from Last Notable Activity

5- Hence overall this model seems to be good.



PRESENTATION TITLE 15