Summary Report

Problem Statement:

With the Data provided by Education X, we have identified the main components of determining whether a lead will convert or not. Also, we have identified the Hot leads on the basis of scores.

Modelling Technique Used:

Generalized Liner Model Regression (Binomial), as it's a case of leads converting or not

Steps Followed for Modelling:

1. Gathering Data
2. EDA - Visualizing Data and Cleaning Data
3. Scaling of Data and Splitting Data into Test and Train Sets
4. Modelling the Data on Train set
5. Predicting the Model on Test Set

Issues Faced:

The main issues faced during the modelling were

1. **Number of Categories in the data**: One of the issues faced was, there were a vast number of categories in the data, and how efficiently we could club them. For now, our strategy was to club all the least data and group them as Others for each column.
2. **Many Imbalanced Columns:** There were many columns where around 90% of the data was a single value. We had to omit such data, to help get a proper accuracy of our model and our model to reduce the Type 2 errors.

Findings from our Model:

Cut Off Probability comes out to be 0.2

Below are the factors contributing to converting a lead from our final model

Generalized Linear Model Regression Results

	coef	std err	z	P>|z|	[0.025	0.975]
const	-2.0888	0.216	-9.654	0.000	-2.513	-1.665
Do Not Email	-1.3012	0.212	-6.134	0.000	-1.717	-0.885
Lead Origin_Lead Add Form	1.0894	0.363	3.001	0.003	0.378	1.801

Lead Source_Welingak Website	3.4138	0.818	4.173	0.000	1.810	5.017
What is your current occupation_Working Professional	1.3403	0.291	4.602	0.000	0.769	1.911
Tags_Busy	3.8040	0.330	11.532	0.000	3.157	4.450
Tags_Closed by Horizzon	7.9562	0.763	10.433	0.000	6.461	9.451
Tags_Lost to EINS	9.1785	0.754	12.177	0.000	7.701	10.656
Tags_Ringing	-1.6947	0.337	-5.036	0.000	-2.354	-1.035
Tags_Will revert after reading the email	3.9665	0.229	17.311	0.000	3.517	4.416
Tags_switched off	-2.2882	0.587	-3.900	0.000	-3.438	-1.138
Lead Quality_Not Sure	-3.3406	0.128	-26.026	0.000	-3.592	-3.089
Lead Quality_Worst	-3.7624	0.850	-4.426	0.000	-5.428	-2.096
Last Notable Activity_SMS Sent	2.7406	0.120	22.847	0.000	2.506	2.976

Please find the Accuracy, Sensitivity, Specificity, Precision and Recall Values of the model below:

Accuracy: 90.67% Precision: 89.4%

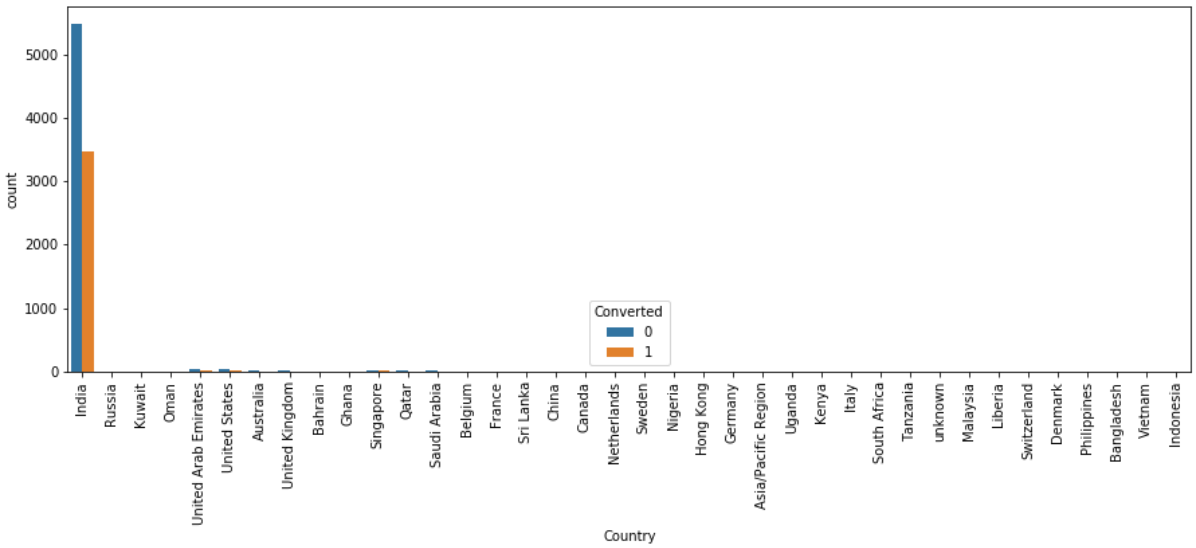
Sensitivity: 84.32% Recall: 84.32%

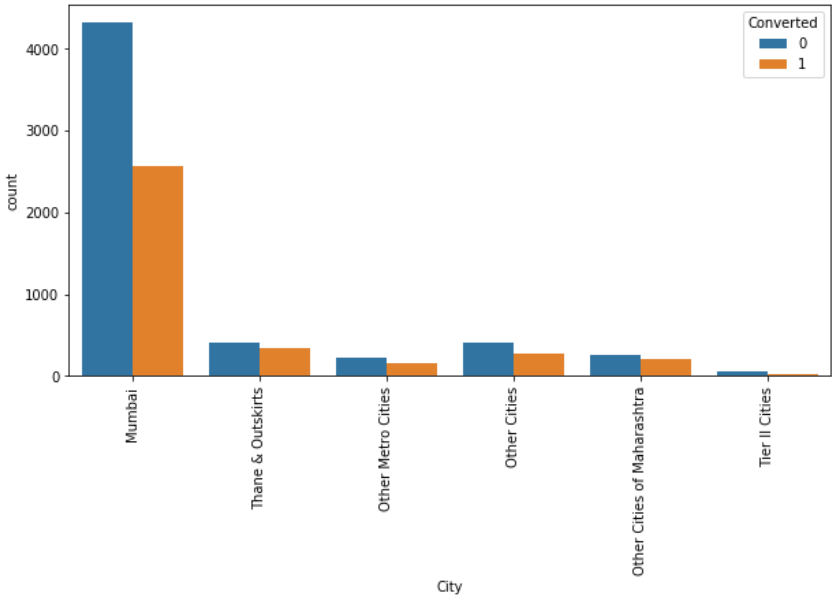
Specificity: 94.3%

Business Aspects of the Model:

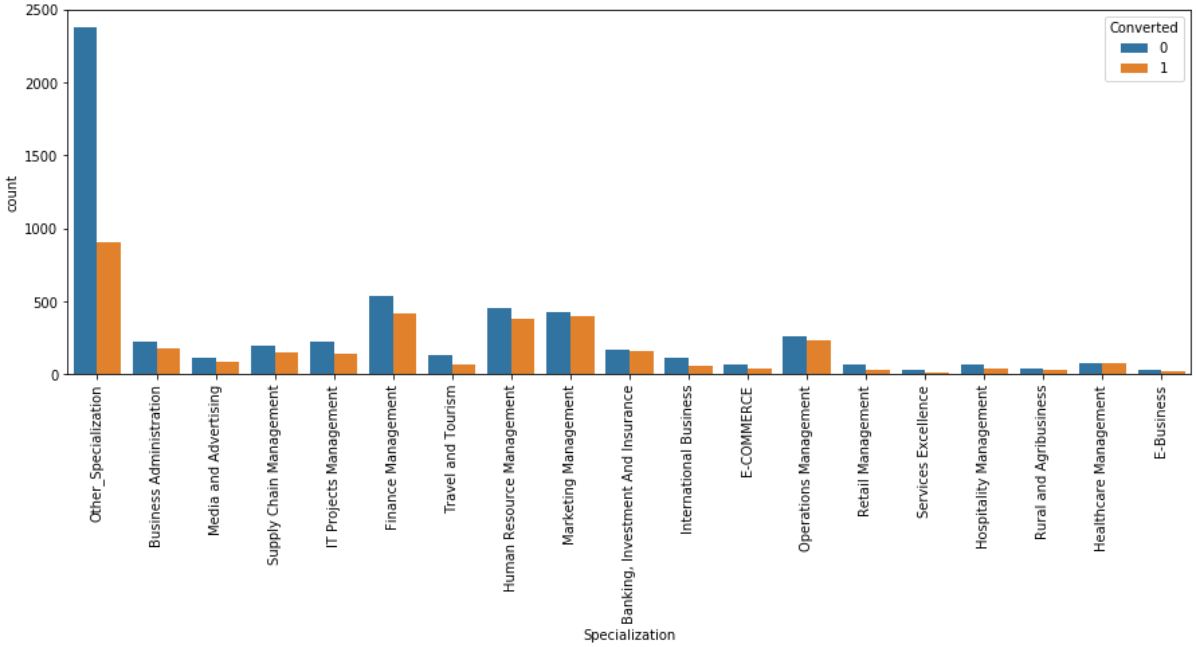
From Business Standpoint, below factors might help in Converting Leads

1. We need to span our Leads from more countries, currently most of our Leads come from India and are in Maharashtra, City. This is seen in Graph below:

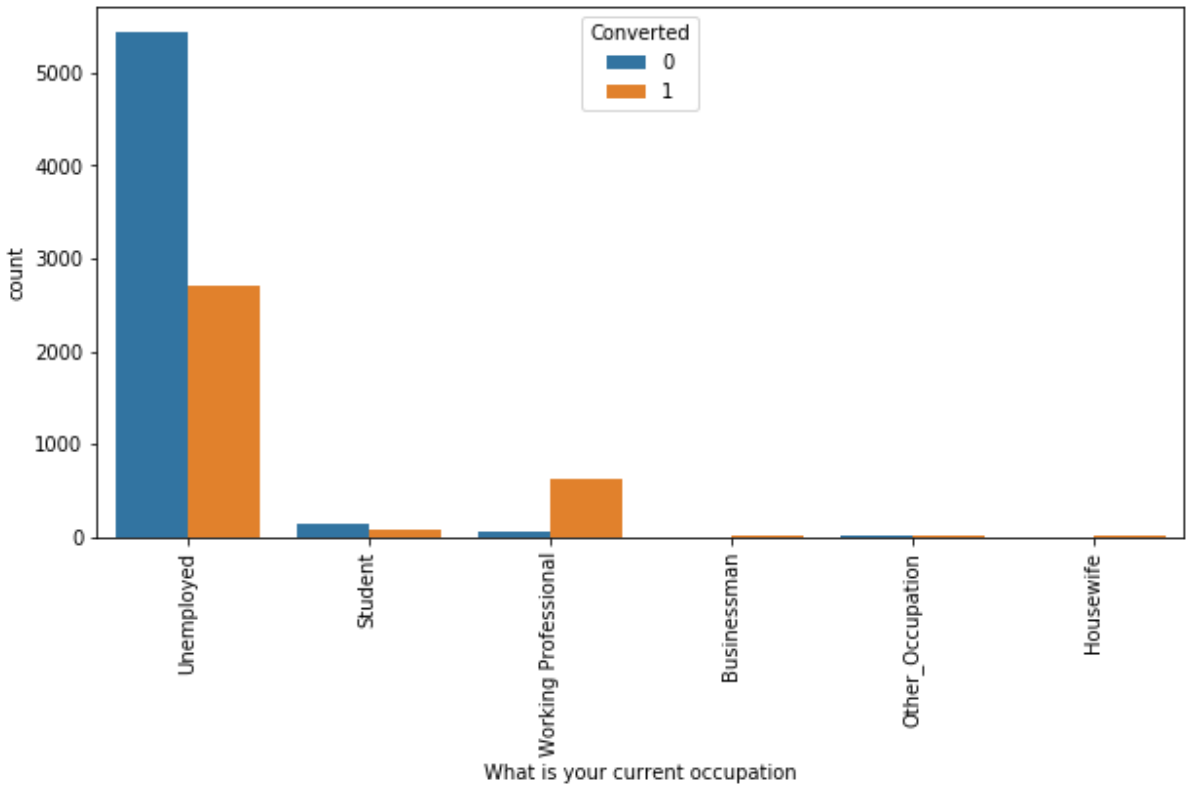




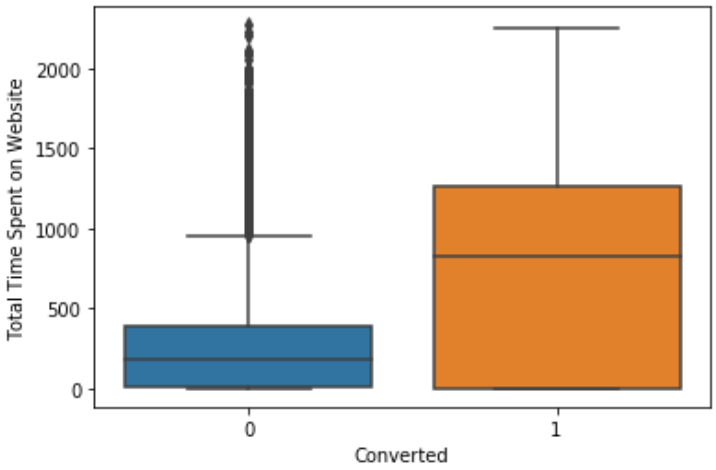
2. We need to include more courses under our repositories, as there are some leads who don't find their relevant courses



3. Since most Leads are looking for Career prospects, we see most leads to be unemployed, we can try to convert them by providing some relaxations or offers in terms of Course Fee



4. We need to improve our websites more engrossing, as more the time the lead spends on the site, more likely he seems to convert and purchase a course



5. We need to get our Leads from Not sure assessed state to High in Relevance