

## Summary

This analysis is done for X Education sells online courses to industry professionals to join their courses. The data provided has lot of information how customers visited the site or time spent on site or forms filled by customers and how they reached X Education organization.

The following steps used:

1. **Libraries:** Import all the needed libraries like pandas, warnings to be ignored, charts to be plotted etc.
2. **Analyze and Data Cleaning:** Import data using libraries and check for null values. There are some values like (**select**) in some columns it's basically empty data, So we can convert all the data into lowercase and the convert into null values using pandas. We notice there are more than one columns has null values. By doing EDA process and kept 30% threshold for cut-off values.
3. **EDA :** There are so huge data from some columns like in country mostly India has having approximately 97% of data and other has the rest.  
Some more like specialization column has not provided data more than the others.  
Most of the converted people has done for career boost experience 95% approximately.  
Most of the users are from Google , Direct traffic lead.
4. **Dummy variables :** The dummy variables are created and used Standard scaler to numeric values.
5. **Train – Test split :** Dataset divided into 70% and 30% for train and test respectively.
6. **Model Building :** We used RFE method to take top 18 features which effect mostly.  
Later we removed highest p-value ( $>0.05$ ) and VIF( $<0.5$ ) respectively.
7. **Model Evaluation :** Confusion matrix made optimum cutoff 0.35 with ROC value calculated accuracy, sensitivity, and specificity of 93%.
8. **Prediction :** Prediction is done on test data with same cutoff value (0.35) and with accuracy, sensitivity, and specificity of 93%.
9. **Precision - Recall :** If you observe precision and recall on test data is not having much difference, 89% and 90%.

I had observed on the data frame there are some columns which are quite interesting like.

Total time spent on website.

Lead source Google.

Last Activity: SMS, Olark chat conversation.

Most of the professionals are opted.

By minimizing the rate of unnecessary phone calls while keeping the focus on new work, the sales team at X Education should be able to effectively manage their time and resources and ensure that they are making the most of this period when they have reached their target for the quarter before the deadline.