Summary

This analysis is done for X Education to sell online courses to industry professionals. The data gave us a lot of insights regarding the potential clients who visit the site, time spent there, how they have reached the site and the conversion rate.

The typical lead conversion rate at X education is around 30% and the company requirement is to build a model where target lead conversion rate is at least 80%.

The following are the steps to build a LOGISTIC REGRESSION:

1. **Importing necessary Warnings and Libraries:** All the libraries and warnings were imported for the implementation.

2. **Reading and Understanding of Data:** This gave us all the insights of the data.

3. **Cleaning of Data:** The data was clean but had null values which were handled by dropping features with more than 40% Missing values. Few NAN values were replaced by “Not Specified” so that much of data is not loose. Checked the Percentage of Rows Missing data. The data was imbalance and was handled.

4. **Exploratory Data Analysis (EDA):** Helps in analysing the data using Univariate and Bivariate analysis which discovers trends, patterns with the help of statistical summary and graphical representations. Through EDA we were able to find out about the outliers present in data and Data imbalance which were handled.

5. **Dummy Variables**: Encodes all of the independent variables as dummy variables allows easy interpretation of the odds ratios, and increases the significance of the coefficients.

6. **Train-Test Split:** Helps to compare our own machine learning model results to machine results. The split was done at 70% and 30% for train and test data respectively.

7. **Scaling Of Data:** Transforms data as it fits within a specific scale. MinMax Scalar is performed.

8. **Model Building:** RFE was done to attain the top 15 relevant variables and model was build checking VIF values and p-value (Keeping the variables VIF < 5 and p-value < 0.05).

9. **Making Predictions**: Confusion matrix was made. We found out that specificity is 96.27% which is good and Senstivity is 88.93% with 0.5 cutoff which needs to be handled.

10. **ROC:** Should be closer to 1 and we are Getting ROC = 0.98 which is very good value which indicates that it is a great model

11. **Probability CutOff:** We concluded that the optimal cut-off for the model is around 0.35 and choosing value to be threshold got decent values of all the metrics – Accuracy (93.01%), Sensitivity (93.18%), and Specificity (92.90%).

12. **Prediction on test data:** With an optimum cut off as 0.35 The Accuracy (93.18%), Recall (93.13%), Specificity (93.22%) and Precession (89.35%).

Conclusion

It seems that Model will predict the Conversion Rate very well.
X Education have a high chance of getting all the potential buyers to buy their courses by focussing on people:

1. Who are visiting the website often and downloading the program sheet.
2. Whose last modified activity falls under the category Modified.
3. Who have opened the Email to have a look over the program.