Lead score assignment Summary

Problem statement

X Education sells online courses to industry professionals. And gets a lot of leads from different sources but the lead conversion rate is very poor. For example, 100 leads come in a day only 30 to 35 gets converted. So, to make this lead conversion more efficient company wants to make some good model to predict the problem and provide some good solutions.

Now, the process followed to make the model and result of the model:

- **Data Cleaning**: first step after loading the data into the system is to check the data is it clean or any null values or outliers is there or not if there is any outliers or null values then it is properly handled. Since there were some data issues so we have made some dummies for that like 'india', 'outside india', 'not provided'.
- **EDA (exploratory data analysis):** a quick EDA is performed after the data cleaning to check the overview of the data by making necessary plots, and heat maps. And if the data looks good and there is no outliers then we can move ahead.
- **Creating Dummies**: the dummy variables were created by using 'get dummies' in the library to make data more efficient. 'min max scaler' was used for scaling the data.
- **Train-test split**: the data was split into two train data and test data in the ratio of 70:30 respectively.
- **Model building**: first of RFE was done to make top 15 important variables. And after RFE one by one irrelevant variables removed using VIF and p-values.
- **Model evaluation**: after the model building it has to be evaluated on the accuracy, sensitivity and specificity which came to be around 80%.
- **Model prediction:** prediction is generally made on test frame data which was done.

After all this process we came to the conclusion that variables which effects the most to potential buyers are:

1. The total time spend on the website.
2. Total number of visits.

3. When the lead source was google, direct traffic, organic search, welingak website.
4. When the last activity was on SMS, olark chat conversation.
5. When the lead origin is lead add format.

After looking at these model and analyzing important features. now X education can make their marketing strategy to increase their revenue and lead into the customer.