Lead Scoring Case Study Summary

The case study focusses on an educational company that wants to increase conversion rates by identifying only the hot leads rather than all leads. To do this, the company uses the logistic regression method to build a model that assigns a lead score of 0 to 100 to each lead, allowing it to target potential customers. A higher number indicates that the lead is hot and hence very likely to convert, while a lower score indicates that the lead is cool and unlikely to do so. We have addressed some additional issues raised by the organization, which the model should be able to adapt to in the event that needs of the business change in the future.

Following steps are used to identify hot leads:

1. **Data Cleaning :**
   We have to remove some columns as it was not required and also replaced null values .

2. **EDA**:
EDA was done to check the condition of the data. We can see that categorical variables were irrelevant and there was no outliers found.

3. **Dummy Variables:**
Dummy variables for categorical variable with 'Select' as option means customer has not selected any value were removed.

4. **Train-Test split:**
The split was done at 70:30 ratio for train and test data respectively.

5. **Model Building:**
RFE was done to obtain the top 15 variables. Rest of the variables were removed manually depending on the VIF values and p-values where variables with VIF<5 and p-value <0.05 were only kept.

6. **Model Evaluation :**
Confusion Matrix was created. Using ROC curve, we found out that accuracy, sensitivity and specificity came to be more than 77%.

7. **Precision-Recall :**
It was used to recheck cutoff of 0.38 and with precision to be 69% and recall to be 77% on test data set.

Variables for the most potential buyers are as follows:

1. Leads who spent more time on website, total number of visits is more likely to convert.
2. People spending higher than average time can be hot leads, so targeting them can be helpful in conversions.
3. When the current occupation is working professional the company has a high chance to get a potential buyer which can buy the course.
4. Maximum lead conversion happened from Landing Page Submission.

5. Sensitivity: 77.08,Specificity: 77.58,Precision: 69.18,Recall: 77.08 which is good indicator that model will work well.
6. We can conclude that the model is in stable state.
7. In business terms, this model has an ability to adjust with the company's requirements in coming future.
8. Important features responsible for good conversion rate or the ones' which contributes more towards the probability of a lead getting converted are :

i) Lead Origin_Lead Add Form
ii) Total Time Spent on Website
iii) What is your current occupation_WorkingProfessional

Thus we can conclude that this model is good but we can't say it perfect but it can be applicable to maximum business scenario to convert leads into customers by rating and identifying the hot leads.