Lead Scoring Case Study Summary

Analysis is done for X Education company and to find out ways to get more industry professionals to buy their course. The data which has been provided tells us about potential customers visiting the website, time spend, how they reached the site and the conversion rate.

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.