Summary

This analysis is done for X Education and to find ways to get more industry professionals to join their courses. The basic data provided gave us a lot of information about how the potential customers visit the site, the time they spend there, how they reached the site and the conversion rate. The following are the steps used:

1. Splitting the Data into Training and Testing Sets. The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
2. Use RFE for Feature Selection. Running RFE with 15 variables as output.
3. Building Model by removing the variable whose p- value is greater than 0.05 and VIF value is greater than 5.
4. Predictions on test data set.
5. Over all 82% accuracy, sensitivity of around 70% and specificity of around 88%

According to the logistics (p-values and VIF), it was found that the factors that mattered the most were:

1. What matters most to you in choosing a course.
2. Country
3. Total Time Spent on Website
4. Lead Origin:
    a) Lead Add Form
5. Last Activity
    a) SMS Sent
    b) Olark Chat Conversation
6. TotalVisits
7. What is your current occupation
8. Last Notable Activity