Summary

This analysis guides X Education to find ways to get more business professionals to take their courses.

Basic data tells us a lot about how people get to places, how long they spend, how they get there, an d the costs. The following are the steps used:

1. Cleaning data :
The data part is clean except for some null values, the selection option does not give us much inform ation so it needs to be replaced with null values. To avoid missing too many objects rare
null values were changed to "NA". Even then these were removed when dummy was made. Since there are more from India and less from outside, change the details to "India", "Outside India" and " Not Provided".

2. EDA:
We did a quick EDA to check the status of the product. It turned out that many items in the categorical variables were not correlated with each other. The results look good
and are not noticeable.

3. Dummy Variables:
A dummy variable is created and then the dummy variable for the "not given" element is deleted. We use MinMaxScaler for arithmetic

4. Train-Test split:

The split was done at 70% and 30% for train and test data respectively.

5. Model Building:
First, RFE was performed to obtain 15 different values. Then, the remaining variables were ma nually removed according to their VIF values and p values
(variables with VIF < 5 and p value < 0.05 were kept).

6. Model Evaluation:
A confusion matrix was constructed. Consensus cut-off values
(using the ROC curve) were then used to determine accuracy, sensitivity and specificity of approximat ely 80%, respectively.

7. Prediction:
Predictions were made on the test data with a visual cutoff of 0.35 and 80% accuracy, sensitivity a nd specificity.

8. Precision – Recall:

This method was also used for identification and a cut-
off value of 0.41 was found with 73% accuracy and 75% recall on test data.