

Summary

This investigation is carried out to assist X education and encourage more business professionals to enroll in their courses. We learned a lot about the potential clients' visitation patterns, length of stay, mode of transportation, and other details from the basic data provided the rate of conversion.

The procedures employed are as follows:

- **Cleaning Data:**

With the exception of a few null values, the data was mostly clean. With a null value, the option select has been replaced because it didn't give us much information.

- **EDA:**

In order to evaluate the quality of our data, we ran a quick EDA. Many components were found to be irrelevant in the category variables. The statistics seem to fall into ranges where no abnormalities have been found.

- **Dummy Variables:**

The variables with "not provided" parts were eliminated once the dummy variables were created. For numerical values, the MinMaxScaler was used.

- **Train-Test Split:**

For the train and test data, the split was performed at 70% and 30%, respectively.

- **Model Building:**

To start, the top 15 relevant variables were determined using REF. After that, the remaining variables were manually removed based on the VIF values and p-values.

- **Model Evaluation:**

A matrix of confusion was created. Afterwards, the ideal cut-off value was employed to determine the 80% accuracy, 80% sensitivity, and 80% specificity of each.

- **Prediction:**

An ideal cut off of 0.35 was used for the prediction, which was performed on the test data frame with 80% accuracy, sensitivity, and specificity.

- **Precision – Recall:**

A cut off of 0.41 was found on the test data frame with a precision of roughly 73.24% and recall of roughly 76.61% using the same process for a recheck.