**Summary:**

The following are the steps used:  
1. Cleaning data:  
Except for a few null values, the data was mostly clean. However, the option chosen had to be changed to a null value because it provided little useful information. To avoid losing a lot of data, a few of the null values were changed to "not provided." Nevertheless, they were later taken out while manufacturing dummies. The elements were altered to "India," "Outside India," and "not provided" because there were a lot of people from India and a small number from elsewhere.  
2. EDA:  
A quick EDA was done to check the condition of our data. It was found that a lot of elements in the categorical variables were irrelevant. The numeric values seem good, and no outliers were found.  
3. Dummy Variables:  
The dummy variables were created and later on the dummies with ‘not provided’ elements were removed. For numeric values, we used the MinMaxScaler.  
4. Train-Test split:  
The split was done at 70% and 30% for train and test data respectively.  
5. Model Building:  
Firstly, RFE was done to attain the top 15 relevant variables. Later the rest of the  
variables were removed manually depending on the VIF values and p-value   
6. Model Evaluation:  
A confusion matrix was made. Later on, the optimum cut-off was used to find the accuracy, sensitivity and specificity which came to be around  
80% each.  
7. Prediction:  
The prediction was done on the test data frame and with an optimum cut of 0.35 with  
accuracy, sensitivity, and specificity of 77%.  
8. Precision – Recall:  
This method was also used to recheck and a cut-off of 0.41 was found with Precision  
around 73% and recall around 75% on the test data frame.