## **Summary**

The model building and prediction is being done for company X Education and to find ways to convert potential users. We will further understand and validate the data to reach a conclusion to target the correct group and increase conversion rate. Let us discuss steps followed:

#### 1. EDA:

- Quick check was done on % of null value and we dropped columns with more than 45% missing values.
- We also saw that the rows with the null value would cost us a lot of data and they were important columns. So, instead we replaced the NaN values with 'not provided'.
- Since India was the most common occurrence among the non-missing values, we imputed all not provided values with India.
- Then we saw the Number of Values for India were quite high (nearly 97% of the Data), so this column was dropped.
- We also worked on numerical variable, outliers and dummy variables.

### 2. Train-Test split & Scaling:

- The split was done at 70% and 30% for train and test data respectively.
- We will Scale the 'Total Time Spent on Website' variables with standard scaler and fitting tranforming the X train dataset

# 3. Model Building

- RFE was used for feature selection.
- Then 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.
- A confusion matrix was created, and overall accuracy was checked.

### 4. Model Evaluation

On Training Data

Accuracy 77.17% Precision 66.6%

Recall 80.4%.

On Test Data

Accuracy 81.52% Precision 75.41% Recall 78.99%

### **CONCLUSION**

TOP VARIABLE CONTRIBUTING TO CONVERSION:

- LEAD NUMBER
- LEAD SCORE
- CONVERTED PROBABILITY

The Model seems to predict the Conversion Rate very well and we should be able to give the Company confidence in making good calls based on this model.