SUMMARY REPORT

This analysis is carried out for X Education to attract more customers to their courses. We learned a lot from the fundamental data on how potential customers use the site, how long they stay there, how they got there, and the conversion rate.

Data cleansing: With a few null values, the data was partially clean. The option select had to be replaced with a null value since it did not give us much information. The columns which had maximum percentage of missing values were removed as it will not help in model building. Also, if the columns are highly skewed with only one category, those columns were dropped. The null values were only partially imputed in order to minimize data loss. Nevertheless, they were later taken out while manufacturing dummies. Many random values in various columns were changed to "Others". Binary variables were converted to 0 & 1.

EDA: Univariate analysis was carried out on both categorical and numerical variables. It was discovered that several of the categorical variables' components were unnecessary. The numerical figures are accurate, and no anomalies were discovered.

Data preparation: The dummy variables were created for categorical variables and for numeric values we used the StandardScaler.

Train-Test split: For train and test data, the split was done at 70% and 30%, respectively.

Feature selection was done using RFE using 15 relevant variables and Optimal Model was determined using Logistic Regression.

Data Modeling & Evaluation: Confusion matrix was created. Later, the accuracy, sensitivity, and specificity were determined using the optimum cut off value (using the ROC curve), which was discovered to be approximately 90% for each.

- On the test data frame, predictions were made using an optimal cut-off of 0.3 and had approximately 90% accuracy, sensitivity, and specificity.
- On the train data set, a cut-off of 0.3 was discovered using this method, with precision around 84% and recall around 89%.

The following factors were discovered to be the most important to prospective customers:

1. Occupation_Unemployed
2. Occupation_Working Professional
3. Last Notable Activity_Email Opened

Recommendations: By connecting with them on various social networks, like Google and others, we can raise the number of working professional leads. We should also focus on improving the conversion rate of unemployed leads in order to increase overall conversion rate. Frequent emails can be sent to people who constantly visit the website to increase conversion.