SUMMARY REPORT:"X EDUCATION LEAD SCORING AND CONVERSION"

Objective:

The analysis is done for X Education, an organization that sells online courses for Industry Professionals. The idea is to market its courses on various platforms and filter the potential buyers of course based on the data provided. The dataset provided to us had various variables which helped identify the Leads who converted/bought the course.

The steps involved in the project were:

DATA CLEANING-

The dataset had null values which were cleaned. Some 'Select' variables were also present which were considered as good as NaN values and therefore removed. Other null values were imputed to a certain variable based on the mode. Some insignificant values were replaced with Other Categories to combine in the categorical values.

EXPLORATORY DATA ANALYSIS (EDA)-

An EDA was done to find the effect of variables on the target variable "Converted". Unnecessary Columns were dropped. Outliers found in the numerical columns were capped when necessary.

DATA PREPARATION-

We created Dummy variables for categorical columns, mapped the binary variables into 1/0 for easy interpretation by Model algorithm.

TRAIN-TEST SPLIT -

The split was done for the final dataset, with 70% data for training the model and 30% data for testing the model.

MODEL BUILDING

Standard Scaler is used and RFE is imported to select 15 features of high importance for building better model. Variables with high p- values (>0.05) and high VIF (>5) were removed gradually.

MODEL EVALUATION-

Confusion matrix was built, ROC curve suggested the cutoff probability of 0.36.

The train dataset measured the accuracy to be almost 80.8%.

The precision and recall of the test data frame came out to be 75%

FEW CONCLUSIONS AND RECOMMENDATIONS-

Conversion Rate: The overall conversion rate of leads is approximately 38.5%.

Lead Origin: API and Landing Page Submission have a large number of leads.

Lead Add Form has a higher conversion rate.

Lead Source: Google and Direct Traffic are major sources of leads. Focus should be on enhancing the google platform and making it more interactive.

Welingak Website has higher conversion rates, which means leads on such platforms are most likely to buy the course. Leads on Olark chat and leads who sought reference are most likely to buy the course.

Leads who visited the site frequently are also the potential leads, so keeping a track of the total visits is important.

Other Specializations should also be considered while filtering potential leads.

Working professionals should be contacted frequently.

Leads tagged as the ones who might revert after reading email could be a potential hot lead