# Summary

For X Education, we conducted the following analysis, which will reveal which industry professionals will enroll in their courses. The customers who visit the website, the amount of time they spend there, how they got there, and the conversion rate are all included in the provided dataset. The accompanying specialized advances are utilized: -

**1. Data Cleaning:**
- Remove the redundant variables/features.
- We have replaced 'Select' with a null value since it did not give us much information.
- Dropped null values having percentage higher than 40
- Handled the missing values by imputing the max related value
- Checked number of unique Categories for all Categorical columns.
- From that Identified the Highly skewed columns and dropped them.
- Treated the missing values by imputing the favorable aggregate function like (Mean, Median, and Mode).
- Detected the Outliers.

**2. Exploratory Data Analysis:**

- We have played out an EDA on our information. It was observed that a great deal of components in the clear cut factors were unimportant. The numerical values appear to be satisfactory, but the outliers were discovered.
- Conducted a univariate analysis of both categorical and continuous variables.
- Performed Bivariate Examination as for Target variable.

**3. Dummy Variables:**

- he dummies variables for each categorical column are created.

**4. Scaling:**

- he data for Continuous variables were scaled using a Standard scalar.

**5. Train-Test Split:**

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

**6. Model Building:**

- utilizing RFE. It lists the most important variables. Later, the variables with a VIF 5 and a p-value of 0.05 were kept, and the irrelevant features were manually removed based on the VIF values and p-value.

**7. Model Evaluation:**

- We have a fair score of accuracy, sensitivity and specificity which is 81%,79% and 83% for test set and 80%, 81%, 80% for train set respectively.

**8. Conclusion:**

- Total time spent on the website: Potential buyers consider the total duration of their visits to the website as an important variable.

-  Number of visits: The absolute count of visits made by potential buyers is still a significant factor.

- Primary source - Olark Chat: When potential buyers interact with the Olark Chat feature as their primary source of engagement, it is considered an important variable.

- Last action - SMS Olark chat conversation: The point at which potential buyers had their last action as an SMS Olark chat conversation is also seen as a relevant factor.