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:

- There was a matrix of confusion. Later on, the best removed regard by using ROC twist was used to find the accuracy, mindfulness and expressness which came to be around 79%.8. Prediction:
- Forecast was finished on the test information outline an ideal cut-off as 0.36 with exactness, awareness and Explicitness of 78%.

8. Conclusion:

We have observed that the potential buyers' most important variables are:

- The entire amount of time spent on the website.
- Absolute number of visits.
- When Olark Chat served as the primary source
- At the point when the last action was SMS Olark talk discussion