**Team 4**

**Summary Of Model**

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. Understand the data and Build the model after cleaning the data and we further validate the data to reach a conclusion to target the correct group and increase conversion rate.

**EDA:**

**Top Variable Contributing to Lead Conversion:**

1. Understood the pattern of dataset by basic info and description of columns.
2. Dropped the columns which have null values more than 40%.
3. Checked for duplicate rows ,also replaced null values of numeric columns with median and replaced null values in categorical columns with mode(high frequent value)
4. Dropping Irrelevant Column
5. Understood the correlation between columns with heatmap and pairplot.
6. Plotting the Countplot using Seaborn Library to check count of top columns in dataset that are going to Contribute more to the Lead Conversion.
7. Dropped top 1% outliers that has been shown in boxplot.
8. Plotted countplot and scatterplot to understand the distribution between pair of important columns.

**Train-Test split& Scaling :**

The split was done at 80% and 20% for train and test data respectively.

We have done Standard scaling on the variables [columns name]

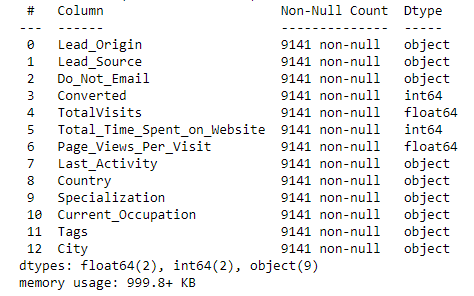
**Model Building 1**

1. Converted categorical data to numeric of various columns using ‘LabelEncoder’
2. Build the Logistic Regression Model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.
3. A confusion matrix was created, and overall accuracy was checked which came out to be **78.07%.**

**Model Building 2**

1. Converted categorical data to numeric of various columns using ‘Dummies’ method.
2. Build the Logistic Regression Model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.
3. A confusion matrix was created, and overall accuracy was checked which came out to be **88.73%.**

**Conclusion**

****

**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 with the help of above columns.**