Team 13:
Summary Report

The goal of this case study was to build a logistic regression model and helps the company find out hot leads and cold leads based on various attributes given in the dataset so they can focus precisely on the customers who are most likely to get converted.

EDA:

1. We also saw that the rows with the "Select'' Value are more and these values should be equally treated as null values since the contribution of this record is equal in both 0 and 1, also no extra information was provided about this record. So, we replace them with a null value to avoid confusion.
2. Clean the data and remove the Null Value
3. Dropping Irrelevant Column
4. We won't be dropping NaN values as they will get automatically taken care of by the get_dummies function in Pandas.
5. Plotting the Heat Map to find the similarity between columns and using Countplot with the help of Seaborn Library to check the count of top columns in the dataset that are going to Contribute more to the Lead Conversion.
6. Converting categorical data into numerical data using the get_dummies function of the Pandas library.

Train-Test split & Scaling :

First, we will apply StandardScaler to all the independent variables.
The split was done at 80% and 20% for train and test data respectively.

Model Building

1. We are building the Logic 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.
2. A confusion matrix was created, and overall accuracy was checked which came out to be 83%.

Evaluating Model based on Classification Report

```
]: from sklearn.metrics import classification_report
   print(classification_report(y_test,lr.predict(x_test)))
```

	precision	recall	f1-score	support
0	0.84	0.89	0.87	1125
1	0.81	0.75	0.78	723
accuracy			0.83	1848
macro avg	0.83	0.82	0.82	1848
weighted avg	0.83	0.83	0.83	1848

With 81% precision, our model is correctly classifying True value on the positive side. And concerning the complete model our model is correctly classifying 75% (recall) of the value on the positive side.

With 84% precision, our model is correctly classifying True value on the negative side. And concerning the complete model our model is correctly classifying 89% (recall) of the value on the negative side.

CONCLUSION

Top Variable Contributing to Lead Conversion:

- **Lead Origin**
- **Do Not Email**
- **Do Not Call**
- **Converted**
- **Total Visits**
- **Total Time Spent on Website**
- **Page Views Per Visit**
- **What is your current occupation**
- **Lead Quality**
- **City**

As given in the problem statement, We successfully built a model with an 80% lead conversion rate was achieved successfully. Also, we narrowed the parameters that affected the conversion and pointed out which attributes to avoid so they can focus on hot leads and utilise manpower productively.