In this Lead Scoring assignment our goal is to our goal is to build a 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.

We can consider that a higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

- In this we will first Importing required Libraries and then Reading and Understanding the Data properly

- After this we will Cleansing the data for EDA And Performing numerical and categorical analysis on the data
- Looking at the data we have some columns like Specialization, Lead Profile, City etc with values as 'Select' where we dont have the info, there are as good as null, so replacing them with Nulls

- Plotting a pairplot for 3 Numerical variables to analysize the trend with Target Converted i.e. Number of leads converted

- We will do the
  ✓ Numerical Inferences:

Total Visits :

Total Time Spent on Website :

Page Views Per Visit :

  ✓ Categorical Inferences:

Lead Origin :

Lead Source :

Do not call or Email;

- Finally we will start to Preparing the data for Model Building
- Converting some binary variables (Yes/No) to 0/1
- Creating dummy features for multi level variables

- Splitting the data into test and train datasets and start for Feature Scaling. From the data, we can see TotalVisits, Total Time Spent on Website and Page Views Per Visit have larger values compared to others. We can normalize the numbers using the StandardScaler method and have all the numbers within small range.

- Checking the Conversion Rate and Looking at Correlations
- Finally Model Building Using all feature variables
- We will create Logistic regression model. Feature Selection Using RFE

Checking metrics using a confusion Matrix, Sensitivity and Specificity

Plotting a ROC curve to check AOC and Finding Optimal Cutoff Point

- Making predictions on the test set

Summary

Based on the coefficient values from model summary, we get top three variables which would affect a lead getting converted into a customer are the following:

1.   Add Form in Lead Origin

2.	Working Professional in What is your current occupation

3.	Last Notable Activity_Modified

To increase the chances of a lead getting converted to a customer, the following variables should be focused upon.

1.	Add Form in Lead Orgin - Of all sources of leads like API's, Searches and Direct traffic, 'Add form' activity was found to be most efficient. Sales teams should focus on leads where forms were submitted from leads.

2.	Working Professional in What is your current occupation - Working professionals are more likely to get converted into customers because of their purchasing parity compared to students or unemployed occupations.

3.	Low in Asymmetrique Activity Index - Leads with low asymmetrique index who have been inactive for a while and have generated some activity generated a lead are probable leads for customers.