Problem Statement:

- An education company named X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google. Although X Education gets a lot of leads, its lead conversion rate is very poor and only about 30% of them are converted.

Goal:

- The Goal is to build a model to identify the Hot Leads and lead conversion rate from Hot leads to be around 80%.

Analysis Approach:

1) Importing Required Libraries, Reading and Inspecting the data set.
2) There are some columns with value as Select and this needs to be replaced with Null values and according Cleaning the Data by checking Null values. Here the Columns *'Lead Number','Country', 'City','Lead Quality','Tags','Asymmetrique Activity Index', 'Asymmetrique Profile Index', 'Asymmetrique Activity Score','Asymmetrique Profile Score','How did you hear about X Education','Lead Profile','Last Notable Activity'"* have more than 30% of null values. So, these columns are removed
3) There are some columns which are not used for analysis such as *'Magazine','Search','Newspaper Article','X Education Forums', 'Newspaper','Digital Advertisement','Prospect ID'*
4) Handling the columns which have null values that are less than 30% like*, Specialization, Lead Source, Last Activity, What matters most to you in choosing a course, Page Views Per Visit .,,etc*
5) Creating Dummy variables for categorical Variables such as *Lead Origin, Lead Source, Last Activity, Specialization, What is your current occupation,etc.*
6) Data Preparation by splitting the data into train and Test Data
7) Scaling the numerical data for train dataset by using Standard Scalar.
8) Model Building using RFE Features
9) Checking the Summary / Accessing the model with Statsmodel
10) Finding all the metrics accuracy, Sensitivity, Specificity and VIF
11) Plotting the ROC Curve
12) Finding the Optimal cut off Point where we get the balanced Sensitivity and Specificity
13) Making the Predictions on the Test Data set
14) Finding all the metrics accuracy, Sensitivity, Specificity values