Summary of Leads Case Study

Aim : To make conversion process more efficient, the company wishes to identify the most potential leads, also known as **'Hot Leads'**.

Accomplishments – The whole assignment is divided into the following steps:
1. <u>Understanding the Data</u> : Understand the attributes of the data. Summarize the data by identifying key characteristics, such as data volume and total number of variables in the data. Understand the problems with the data, such as missing values, inaccuracies, and outliers.
2. <u>Data Cleaning and Visualisation</u> : Identifying the *incorrect*, *incomplete*, *inaccurate*, *irrelevant* or *missing* part of the data and then modifying, replacing or deleting them according to the necessity.
   The graphical representation of information and data is called visualisation. All the features are visualised using countplot, boxplots, bargraphs etc.
3. <u>Exploratory Data Analysis</u> : Exploratory Data Analysis is a crucial step before you jump to machine learning or modelling of your data. It provides the context needed to develop an appropriate model – and interpret the results correctly.
4. <u>Data Preparation</u> : Data preparation (also referred to as "data pre-processing") is the process of transforming raw data so that data scientists and analysts can run it through machine learning algorithms to uncover insights or make predictions.
5. <u>Choose a Model & Train the Model</u> : The process of training an ML model involves providing an ML algorithm (that is, the learning algorithm) with training data to learn from. The term ML model refers to the model artifact that is created by the training process.
6. <u>Model Evaluation</u> : Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future.
7. <u>Make Predictions</u> : Prediction" refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome, such as whether or not a customer will be converted to take the course.
8. <u>Result of analysis</u> : The sensititvity turns out to be approx 85% which is much greater than the initial (30%). Hence it is good to say that the aim is achieved and selected features are obtained which contribute most to enhance probability.

Some Inferences :
- Converting 'Select' values to NaN.
- Replacing null values in categorical features with most occurring value and small categories to "Others".
- Based on the univariate analysis we have seen that many columns are not adding any information to the model, hence we can drop them for further analysis. Like 'Lead Number', 'What matters most to you in choosing a course', 'Search', 'Magazine', 'Newspaper Article' etc.
- Converting some binary variables (Yes/No) to 1/0.
- Calculating accuracy (~92%) and sensitivity (~85%)