# LEAD SCORE CASE STUDY SUMMARY

Below are the steps followed for building the logistic regression model for lead score

1. **Data cleaning:**
    - Analyzed the data by using predefined methods info () and describe ()
    - Removed redundant columns and replaced 'Select' variable with NAN. As select means customer has not answered to the question.
    - Removed the columns which are having more that 30% null values
    - Handled other missing values by replacing the values with most frequently used value for their respective columns
    - Performed EDA on univariate and bivariate variables

2. **Data Preparation:**
    - Converted all the YES-NO values in the dataset to binary values
    - Created dummy variables for the multicategory column values and dropped the columns for which dummies are created

3. **Test-Train data split and Model Building:**
    - Split the data into test-train data and scaled the dataset using StandardScaler
    - As the number of columns are more, initially performed Recursive feature elimination (RFE) with rfe count 15
    - Performed manual feature elimination and removed the features having high p-value
    - Also, used variance-inflation factor (VIF) to delete the features having with VIF greater than 5(None of the features were having more than 5 VID for the dataset)
    - Plotted ROC curve to determine the cutoff
    - We checked the precision and recall with accuracy, sensitivity and specificity for our final model and the tradeoffs
    - We have given lead score to the test dataset for indication that high lead score are hot leads and low lead score are not hot leads.

4. **Conclusion:**

    - The Accuracy, Precision and Recall score we got from test-train set acceptable in range.
    - We have high recall score than precision score which we were exactly looking for.
    - In business terms, this model has an ability to adjust with the company's requirements in coming future.
    - This concludes that the model is in stable state.

- Important features responsible for good conversion rate or the ones' which contributes more towards the probability of a lead getting converted are:
    a. Total Time spent on website
    b. Lead Origin_Lead Add Form and
    c. When Last Notable Activity – Had a phone conversation