SUMMARY REPORT

A case study of 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 . 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.

Steps to follow

- Data Exploration
- Data Cleaning
- Data Preparation
- Variable Selection
- Modelling
- Dropping of variable
- Evaluation : Evaluation Metrix: Confusion Matrix, Recall/Sensitivity/F1-Sure matrix is used to check the performance of the model.

After splitting the data into train and test data sets and the multiple columns which doesn't add value to data are dropped by RFE technique than by p-value and VIF value.
Columns having high p value and VIF values are dropped.
Keep the variable with
P value<0.05
VIF value< 5
Three variables which are important for lead generation are : Time spent on website , lead source and Tags.
After dropping the variable construct confusion matrix to find out accuracy,sensitivity and specificity to check the performance of model.