**LEAD SCORING CASE STUDY**

1. A brief summary report in 500 words explaining how you proceeded with the assignment and the learnings that you gathered.

**Answer**:

So here are the following summary report of the assignment that we have gone through and how we achieved this milestone.

1. Data Loading
2. We first Imported all necessary Libraries.
3. We loaded data using csv
4. Data Cleaning
5. We first checked is there any null values in percentage and after that we dropped those columns whose null values was >40%
6. We cleaned those variables where they contain redundant variables.
7. Now after that we found some columns having label as ‘select’ which means customer has not chosen to answer. So we took those values and named it has null values.
8. And after that we made some progress for those columns with imputed values with maximum number of occurrences for a column. We got to know the exact value counts using EDA countplot for various variables.
9. Data Transformation
10. Changed the Multicategory labels into dummy variables and binary variables to ‘0’ and ‘1’.
11. We also mapped some Categorical Variables to [yes:1, no:0]
12. After that we made some outliers for Numeric variables which helped us to find outliers
13. Data Preparation
14. Split the dataset into Train and Test Dataset and scaled the dataset
15. We found some correlations which were highly correlated so hence we dropped them.
16. Model Building
17. We created RFE model with 20 variables and trained some columns with support and rank
18. For our model we created 9 models with the help of statsmodels and kept on making the model till we get p-values 0 and VIF value less than 5.
19. Whose p-values were high we dropped those columns so that it wont effect our final model.
20. All these models we checked the optimal probability cutoff by finding points and checked the accuracy, sensitivity and specificity.
21. We checked the precision and recall with accuracy, sensitivity and specificity for our final model
22. Prediction made on Train set.
23. We created new columns and predicted those variable whose values were >0.5.
24. We did model evaluation on the test set like checking the accuracy, recall to find how the model is.
25. After that we found the score of accuracy and sensitivity from our final test model is acceptable.
26. 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.
27. Summary

Learning gathered are below:

1. Test set is having accuracy, recall/sensitivity in an acceptable range.
2. We can conclude that our model is having stability an accuracy with adaptive environment skills. So we can easily predict for the upcoming future through this model
3. So following are the responsible for good conversion rate

* Total Visits
* Page Views Per Visit
* Total Time spent on website
* Lead Origin Lead add form