Leads Scoring Case Study

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

Answer: Below are the steps how we have proceeded with our assignments:

1. **Inspecting the Dataframe:**
   a. Checking data type and getting information on data
   b. Looking at the statistical aspects of the data-frame.
2. **Data cleaning**
   we found that some columns are having label as 'Select' which means the customer has chosen not toanswer this question. The ideal value to replace this label would benull value as the customer has not opted any option. Hence, we changed those labels from 'Select' to null values.
   a. Removed columns having more than 40% null values
   b. For remaining missing values, we have imputed values withmaximum number of occurrences for a column.
   c. We have normalise some column also.
3. **Data Transformation:**
   a. Changed the multicategory labels into dummy variables and binaryvariables into '0' and '1'.
   b. Checked the outliers and capping at 99 percentiles.
   c. Removed all the redundant.
4. **Data Preparation:**
   a. Splited the dataset into train and test dataset and scaled the dataset.
   b. After this, we plot a heatmap to check the correlations among thevariables.
   c. Found some correlations and they were dropped.
5. **Model Building:**
   a. We created our model with rfe count 19 and 15 and compared the model evaluation score like AUC and choose our final model with rfe19 variables as has more stability and accuracy than the other.
   b. For our final model we checked the optimal probability cutoff by finding points and checking the accuracy, sensitivity and specificity.
   c. We found one convergent points and we chose that point for cutoffand predicted our final outcomes.
   d. We checked the precision and recall with accuracy, sensitivity andspecificity for our final model and the tradeoffs**.**
   e. Prediction made now in test set and predicted value was recoded.
   f. We did model evaluation on the test set like checking the accuracy, recall/sensitivity to find how the model is
   g. We found the score of accuracy and sensitivity from our final testmodel is in acceptable range.
   h. We have given lead score to the test dataset for indication that highlead score are hot leads and low lead score are not hot leads.

6. Conclusion:

Learning gathered are below:

a. Test set is having accuracy, recall/sensitivity in an acceptablerange.
b. In business terms, our model is having stability an accuracy with adaptive environment skills. Means it will adjust with thecompany's requirement changes made in coming future.
c. Top features for good conversion rate:
    1. **Last Notable Activity_Had a Phone Conversation**
    2. **Lead Origin_Lead Add Form**
    3. **What is your current occupation_Working Professional**