Lead Scoring Case Study

1. **Data Cleaning:**
    a. There are some columns which have “Select” which means the customers has chosen not to answer the question. We find the columns and replace the “Select” with null value.
    b. We then replace the null value with median for categorical values and mode for numeric value.
    c. We find the column that has same values but in different format(capital letter and small letter). We fix the issue by making the label into a same format.
2. **Data Transformation:**
    a. Create the dummy variables for multicategory labels.
    b. We convert “Yes” and “No” with 1 and 0.
    c. Remove the redundant and repeated columns.
3. **Data Preparation:**
    a. Split the dataset into train and test dataset and normalize the dataset.
    b. We check the correlation of the variables and create a heatmap.
    c. We found a lot of variables that are correlated and that will be dropped during RFE.
4. **Model Building:**
    a. We create a model and perform RFE with 15 variables.
    b. For our model we check the optimal probability cutoff by checking the accuracy, sensitivity and specificity.
    c. We found one convergent points and we chose that point for cutoff and predicted our final outcomes.
    d. Check the precision and recall with accuracy, sensitivity and specificity for our final model and the tradeoffs.
    e. Prediction made now in test set and predicted value was recorded.
    f. We do model evaluation on the test set like checking the accuracy, recall/sensitivity of the model.
    g. We found the score of accuracy and sensitivity from our final test model is in acceptable range.
    h. We give lead score to the test dataset for indication that high lead score are hot leads.

5. **Conclusion:**
    a. Test set is having accuracy, recall/sensitivity in an acceptable range.
    b. In business terms, our model us having stability an accuracy to adapt any future changes to the company's requirements.
    c. Top features for good conversion rate are:
        i. Closed by Horizzon
        ii. Lost to EINS
        iii. Will revert after reading the email