SUMMARY OF LEADS SCORING ASSIGNMENT

DATA READING

- Checking the shape, info, description of the data.

CLEANING THE DATA

- Checking the data and removing useless columns.
- Converting the label "select" present in columns to null values.
- Then checking for null values and removing columns with 35% null values.
- Columns with less missing values are imputed with most occurring label in that column.

DATA TRANSFORMING

- Changing labels to dummy variables and "YES" to 1 and "NO" to 0.
- Removing the duplicate columns.

TEST-TRAIN SPLIT

- Splitting the data in test-train and scaling the data.
- Plotting the heat map for checking correlation.

MODEL BUILDING

- Running ref on 15 variable and checking the linear model regression results.
- Removing the columns with high P value one by one to and checking the VIF value.
- Predicting on the train model.
- Checking the ROC curve and checking the accuracy, specificity and senility.

- Checking and calculating precision and recall with cut-off of 0.35 and 0.41 on train and test data.

CONCLUSION

- We can say that for good conversion rate which leads getting converted are:

- Tags_Lost to EINS

- Tags_Closed by Horizzon

- Tags_Will revert after reading the email