SUMMARY

This Case Study on Lead Score is related to a Education Company: The X Education

X Education Company provides online courses for working Professionals, here the company needs help to enhance its sales that is the leads that are most likely to get converted into paying customers.

The company has provided us a data set from which it expects us to build a model wherein the lead score needs to be allotted to each if the leads such that the customers with higher lead score have higher chances of getting converted and customers with lower lead score have lower chances of getting converted.

We were provided with 37 fields/area where we needed to focus on and get the optimum result and help the company reaching its target of 80% conversion which is only 30% now

STEPS FOLLOWED:

1. **Data Understanding**
   Here we went through the dataset for leads .
2. **Data Cleaning**
   Basic data cleaning has been done , variables that were having more than 40% null values has been dropped
3. **Data Preparation**
   a. We went through the null/missing values whose absence might affect the analysis and imputed them with median, outliers were also taken care of.
   b. Dummy variables were created for the categorical variable

4. **Exploratory Data Analysis**
   - We did Univariate analysis, plotted few important features with target convert variable.
   - Performed Train Test Split on the dataset with 70% trained data and 30% test data
   - Checked for Correlation between each variables using heatmap

5. **Model Building**
   - We tried to evaluate important features for our model using RFE.
   - After that we created stats using model and validated the p value

- Checked for p value which has to be maintained at p<0.05 and VIF score <5
- VIF score determined the correlation between each independent variable
- The above conditioned were examined by dropping single features at a time based on p-value and VIF value
-

6. **Model Evaluation**
    - We determined the threshold from the ROC curve
    - Evaluated accuracy, Sensitivity, Specificity, precision, recall
    - Focused on making the model 80% accurate which the company wanted

7. **Predictions on Test Data set**

Accuracy: 93.65 %

Specificity: 93.85 %

Sensitivity: 93.34 %

CONCLUSION

- Tags_lost to EINS, Tags_Closed by Horizon, Lead_Source_Welingak Website are the features that contribute most towards getting the proper probability of a lead getting converted. So company needs to focus on this
- Accuracy of the model is 93.65%with high sensitivity and specificity that will help the company reaching their target customers