Summary – Lead Score Case Study

Importing & Inspecting Data:

- Importing all the required libraries
- Importing Data in Data Frame and understanding the data was a critical process.

Data Cleaning and Preparation:

- After understanding the data, we needed to check the columns with the amount of null or NA Values in them.
- We decided to drop columns which had more than 30% of Null Values
- We also dropped the columns which we did not require for analysis.
- Then we imputed the null values in all the columns.

Data Analysis:

- We created various charts & combinations of results to derive meaningful conclusions.
- We handled the Outliers in columns.
- We created Dummy Variables.
- We did the Test-Train Split for a better model.
- We tested Feature Scaling and
- Looked into Correlations

Model Building:

- We build our model with the help of RFE with 15 variables.
- Assessed the model.
- We then removed the insignificant variables using the P-Value Score.
- After the P-value check all our variables had a good value of VIF and there was no multicollinearity. Hence, we need not to drop any more variables and we could proceed with making predictions using this model only.
- We created new column 'Predicted' with 1 if Converted_Prob > 0.5 else 0.
- Checked the Metrics (Confusion Matrix, Accuracy, Sensitivity,

Specificity, False Positive Rate, Positive Predictive Value and Negative Predictive Value).

- We did model evaluation on the test set by checking the Accuracy, Sensitivity, Specificity, False Positive Rate, Positive Predictive Value and Negative predictive value to find how the model is.
- After all the tests & evaluations we concluded that the model built, and the scores achieved on our final modelwere in the acceptable range.

Conclusion:

- The Accuracy, Precision and Recall score we got from the test data are in the acceptable region.
- Accuracy, Sensitivity and Specificity values of test set are around 76%, 76% and 77% which are approximately closer to the respective values calculated using trained set.
- Also, the lead score calculated in the trained set of data shows the conversion rate on the final predicted model is close to 80% (i.e., 78%)
- Hence overall this model seems to be good.
- A customer Lead sourced by "Welingak Website" is a Hot Lead.
- A customer who is currently "Working Professional" or "Unemployed" is a Hot Lead.
- Total Time Spent on Website gets high conversion.