Lead Scoring Assignment- Summary

1. We have used the logistic Regression model to solve the business problem in hand, Since it is a binary classification problem we have used the logistic regression model
2. We have followed the below standard steps to solve the problem
   - Read and inspect the data
   - Clean the data- Null values, not useful features, and considered Select data point as Null value and removed the data points with high null values
   - Data Processing- Dummy variable creation for categorical variables, used standard scaling for numerical columns and also map the binary yes/no with 1 and 0
   - Train test split
   - Model building- used RFE for choosing top 20 features
   - Used p values and VIF scores further eliminate the variables
   - Use the model for prediction on train data
   - Model evaluation based on some metrics like accuracy, sensitivity, specificity etc
   - Model assessment on test data
   - We have also used ROC to find the thresh hold probability
3. Run the multiple models and finally selected Model 3 as the best performing model based on the evaluation metrics
4. Answered the questions of the business team based on the model