Problem Statement

Prosthetic Systems is a medical health provider with value added services towards its customer. Due to tough completion in the niche market, its struggling to sell its products and make profits. As they deal with only specialized products having very low volume, there is no way they can increase the sales to make up the revenue and the profit. The management is deciding to have a separate vertical that can finance the customers for the sales and services for these premium products. This will enable them to stay in the market, as well as they can expand their finance division if that rolls out well. As the company does not have a full banking experience or is not supervised by a national or international banking regulatory agency so they are going to tie up with a investment firm to start the company with some shared revenue model. Now the challenge for them is to determine how they are going to select the customers to reach out. The company wants to build a model that will help them identify the potential customers who have a higher probability of purchasing the loan. This will increase the success ratio while at the same time reduce the cost of the campaign.

Goal

You are brought in as a consultant and your job is to build the best model which can classify the right customers who have a higher probability of purchasing the loan. You are expected to do the following:

- EDA of the data available. Showcase the results using appropriate graphs
- Build appropriate models on both the test and train data (CART, Random Forest). Interpret all the model outputs and do the necessary modifications wherever eligible (such as pruning)
- Check the performance of all the models that you have built (test and train). Use all the model performance measures you have learned so far. Share your remarks on which model performs the best.

The solution should be a report with proper explanation, Python code, interpretation & Conclusion.

Evaluation:

You have to submit 2 files:

1. Business Report not exceeding 2000 words.
   In this you need to submit all the answers to all the questions in a sequential manner. Your answer should include detailed explanations & inferences to all the questions. Your report should not be filled with codes. You will be evaluated based on the business report.

   Sequence:
   - **a.** EDA -Basic data summary, Univariate, Bivariate analysis, graphs
   - **b.** Applying Supervised Machine Learning Techniques (Test & Train)
     i. Applying CART <plot the tree>
     ii. Interpret the CART model output <pruning, remarks on pruning, <plot the pruned tree>
     iii. Applying Random Forests<plot the tree>
     iv. Interpret the RF model output <with remarks, making it meaningful for everybody>
   - **c.** Various Model Performance Measures (Test & Train) :
     i. Confusion Matrix
     ii. Remarks on Model validation exercise <Which model performed the best>
     iii. Build the Model using the algorithm selected on the last step and Interpret the results
2. Python code file : You need to submit the work using **Github** and provide the **link** in your report. This is a must and will be used for reference while evaluating.

Notes

- Any assignment found copied/ plagiarized will not be graded and marked as zero.
- Please ensure timely submission as post deadline assignment will not be accepted.
- Please ensure to upload the correct file. We will not entertain any request for file re-submission after the project due date.