1. Objective: Build Decision Tree models using C5.0 and rpart

Steps:

1. Load Data into R(Use CustomerData.csv)
2. Data preparation
   a. Check the structure and summary of the dataset
   b. Check for missing values
   c. Convert necessary columns into factors

3. Split the Data into test and train using "createDataPartition" from caret package.

4. Build a regression model using" rpart" for "TotalRevenueGenerated" .

   a) Build the model and understand the "cp" value

   ```
   DT_rpart_Reg<-rpart(TotalRevenueGenerated~.,data=train,method="anova")

   DT_rpart_Reg<-rpart(TotalRevenueGenerated~.,data=train,method="anova",control =
   rpart.control(cp = 0.001))
   printcp(DT_rpart_Reg)
   ```

   b) Calculate the evaluation metrics

   ```
   predCartTrain=predict(DT_rpart_Reg, newdata=train, type="vector")
   predCartTest=predict(DT_rpart_Reg, newdata=test, type="vector")

   regr.eval(train[,"TotalRevenueGenerated"], predCartTrain)
   regr.eval(test[,"TotalRevenueGenerated"], predCartTest)
   ```

5. Build a classification model model using" rpart" and "C5.0".
   1. Tweek the data set for classification by binning the target "TotalRevenueGenerated"

      If TotalRevenueGenerated < 150 ," Regular " , else, "Premium"

   2. Build Classification model using "C5.0"

      ```
      library(C50)
      DT_C50 <- C5.0(Revenue~.,data=train)
      summary(DT_C50)
      ```

   3. Calculate the evaluation metrics

      ```
      pred_Train = predict(DT_C50,newdata=train, type="class")
      pred_Test = predict(DT_C50, newdata=test, type="class")

      confusionMatrix(train$Revenue,pred_Train)
      confusionMatrix(test$Revenue,pred_Test)
      ```

   4. Repeat the above experiment using "rpart" model and tune for "cp" value.