# Read in the red dataset

data1<-read.csv("~/wine\_red.csv")

# Look at the column names

names(data1)

#data1<- na.omit(data1)

# Split the dataset into a training and test dataset using a split of 75% training

# and 25% test

set.seed(42)

nrow(data1)

trainingObs\_red<-sample(nrow(data1),0.75\*nrow(data1),replace=FALSE)

# Create the training dataset

trainingDS\_red<-data[trainingObs\_red,]

View(trainingDS\_red)

# Create the test dataset

testDS\_red<-data[-trainingObs\_red,]

View(testDS\_red)

write.csv(testDS\_red,file="red\_test.csv")

#smote training data

table(trainingDS\_red$BQ2\_quality)

smoted\_red<-SMOTE(BQ2\_quality ~ ., trainingDS\_red, perc.over=400, perc.under= 125)

table(smoted\_red$BQ2\_quality)

write.csv(smoted\_red, file="red\_train\_smote.csv")