1. Use the given link below and locate the bank marketing dataset. Data Set Link

Perform the below operations:

a. Is there any association between Job and default?

p\_duration <- ggplot(mydata, aes(factor(y), duration)) + geom\_boxplot(aes(fill = factor(y)))

b. Is there any significant difference in duration of last call between people having housing loan or not?

for(level in unique(mydata$job)){

mydata[paste("job", level, sep = "\_")] <- ifelse(mydata$job == level, 1, 0)

}

c. Is there any association between consumer price index and consumer?

set.seed(1)

training\_size <- floor(0.80 \* nrow(mydata))

train\_ind <- sample(seq\_len(nrow(mydata)), size = training\_size)

training <- mydata[train\_ind, ]

testing <- mydata[-train\_ind, ]

preProcValues <- preProcess(training, method = c("center", "scale"))

scaled.training <- predict(preProcValues, training)

scaled.testing <- predict(preProcValues, testing)

d. Is the employment variation rate consistent across job types?

down\_training <- downSample(x = scaled.training[, -ncol(scaled.training)], y = scaled.training$Class)

up\_training <- upSample(x = scaled.training[, -ncol(scaled.training)], y = scaled.training$Class)

smote\_training <- SMOTE(Class~., data = scaled.training)

rose\_training <- ROSE(Class~., data = scaled.training, seed=2)$data

e. Is the employment variation rate same across education?

ctrl <- trainControl(method = "repeatedcv", repeats = 5,

classProbs = TRUE,

summaryFunction = twoClassSummary)

orig\_fit <- train(Class~., data = training,

method = "rpart",

metric = "ROC",

trControl = ctrl)

f. Which group is more confident?

set.seed(1)

training\_size <- floor(0.80 \* nrow(mydata))

train\_ind <- sample(seq\_len(nrow(mydata)), size = training\_size)

training <- mydata[train\_ind, ]

testing <- mydata[-train\_ind, ]