```
require(ggplot2)

require(boot)

require(ROCR)

require(OptimalCutpoints)

require(caret)

require(plyr)

require(rpart)

require(rattle)

require(reshape2)


require(odbcConnect)

db <- odbcConnect("DatabaseName")

f <- sqlQuery("select * from table", stringsAsFactors=FALSE)

f <- read.csv("~/table.csv", stringsAsFactors=FALSE)


f <- f[complete.cases(f),]


f$additionalUsers <- with(f, count_admins + count_nonadmins)


vars <- c("var1","var2","var3","var4","var5","var6","var7","var8") tierSeg <- aggregate(f[vars],
by=list(f$pdStat), mean)


colnames(tierSeg)[1] <- "IsPaid"

tiers.m <- melt(tierSeg, id.vars='IsPaid')

ggplot(data=tiers.m, aes(IsPaid, value)) + geom_bar(aes(fill = IsPaid), stat="identity", position = "dodge")
+ facet_wrap(~ variable, nrow=2, scales="free_y")


m <- rpart(paid ~ var1 + var2 + var3 + var4 + var5 + var6 + var7 + var8, data=f, method="class")
```

```
fancyRpartPlot(m)


a<-data.frame(m$variable.importance)

a$variable <- factor(rownames(a), levels=rownames(a))

x <- transform(a, variable=reorder(variable, -variable.importance) )

ggplot(data=a, aes(x=variable, y=importance)) + geom_bar(stat="identity", fill="lightblue")


f$prob <- predict(m, f)[,"TRUE"]


pred <- prediction(f$prob, f$paid) perf <- performance(pred, measure="sens", x.measure="ppv")
plot(perf, col=rainbow(10))


optimal.cutpoints(X="prob", status="paid", tag.healthy=TRUE, methods=c("MaxKappa"), data=feat,
direction=">", control=cont) f$predict <- f$prod> .2


  Reference

Prediction FALSE TRUE

 FALSE    1060  62

 TRUE     73    202


 Accuracy : 0.9207672

 95% CI : (0.9478534, 0.971242)

 No Information Rate : 0.9442023

 P-Value [Acc > NIR] : 0.006509636


 Kappa : 0.7303789

 Mcnemar's Test P-Value : 1.000000000


 Sensitivity : 0.6512676056
```

Specificity : 0.95876270

Pos Pred Value : 0.6745454545

Neg Pred Value : 0.93966728