**Sas hw3**

109306045資管二黃筠茜

library(datasets)

library(tidyverse)

library(magrittr)

library(cluster)

library(NbClust)

library(factoextra)

#data(train)

#1

library(MASS)

trainI <- sample(1:352, 75)

traind <- train[trainI,]

testd <- train[-trainI,]

##Decision tree

install.packages("rpart")

install.packages("rpart.plot")

library(rpart)

tree <- rpart(satisfaction ~. ,data=traind, method="class") #inside part: y~X1+X2+...

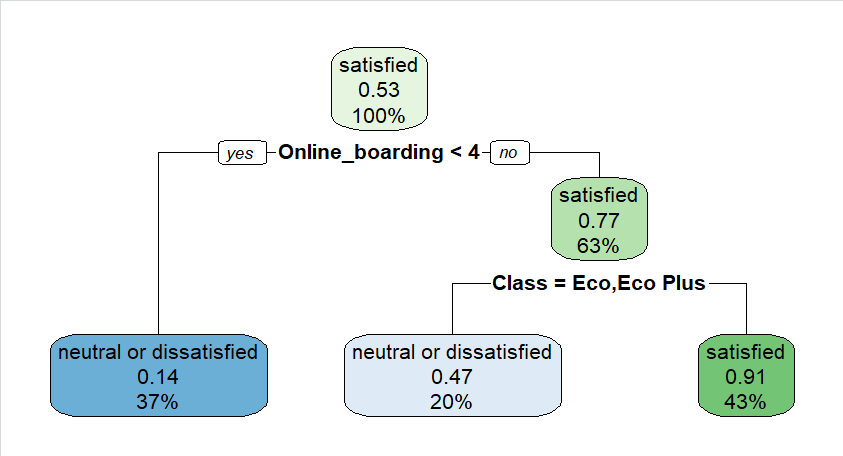
pred <- predict(tree, newdata=testd, type="class")

table(Real = testd$satisfaction, Predict = pred)

library(rpart.plot)

rpart.plot(tree)

rpart.rules(tree,cover=T)



由此圖可看出satisfied neutral以及dissatisfied的預測

### Random Forest ###

install.packages("randomForest")

library(randomForest)

rf <- randomForest(satisfaction ~ ., data = train, importance=TRUE, ntree=100)

importance(rf)

#2

newdata <- train %>%

mutate(

#type = ifelse(TypeofTravel=='Business travel',1,0),

#age = (train$Age - min(train$Age)) / (max(train$Age) - min(train$Age)),

online.boarding=((train$Online\_boarding) - min(train$Online\_boarding)) / (max(train$Online\_boarding) - min(train$Online\_boarding)),

gender = ifelse(train$Gender=='Male',1,0)

)

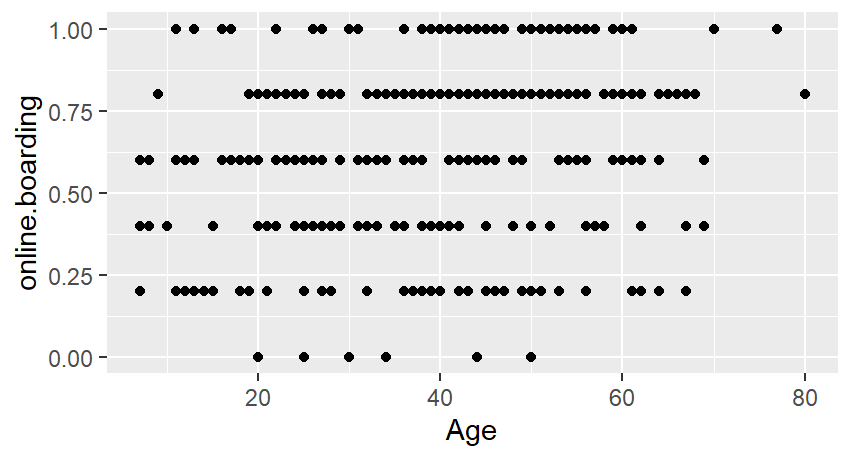
view(newdata)

newdata<-newdata[,-c(1,2,3,

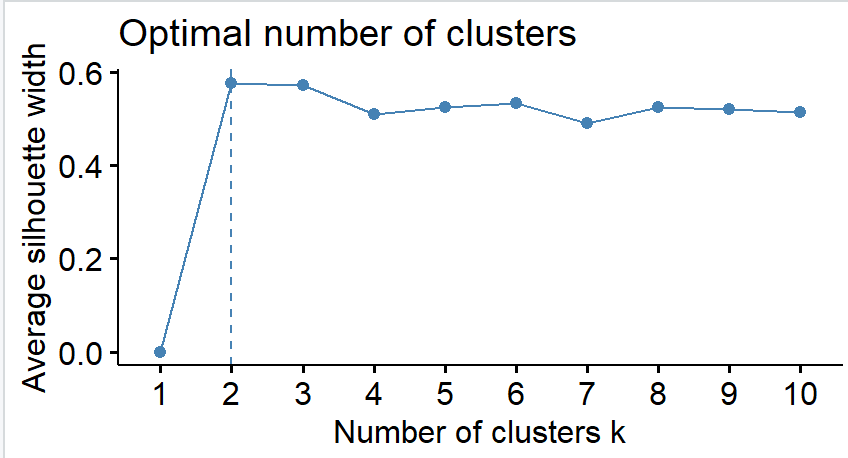
4,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25)]

ggplot(newdata, aes(x=Age, y=online.boarding)) +

geom\_point()



Online boarding與age的關係圖

fviz\_nbclust(newdata[,1:2], FUN = kmeans, method = "silhouette")

km <-

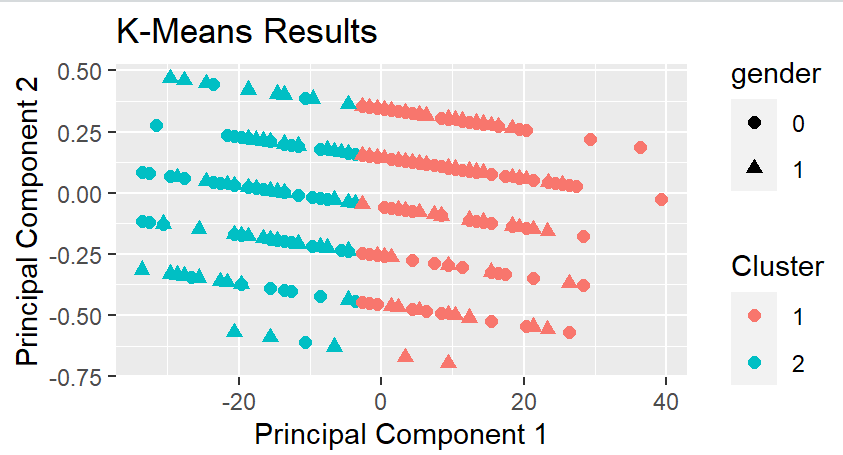
可看出適合的k是二

kmeans(newdata[,1:2], centers=2,nstart=20)

install.packages("useful")

library(useful)

plot(km, data=newdata, class="gender")



可看出分組

fviz\_cluster(km,

data = newdata[,1:2],

geom = c("point"),

ellipse.type = "norm")

cc = km$cluster

data = cbind(data,cc)

ggplot(data, aes(x=Age, y=online.boarding,color=as.factor(cc))) +

geom\_point()



#可以從圖表中看出年輕乘客對線上服務較不滿意,如果能改善介面跟功能貼近年輕族群應能提升年輕群的滿意度

#年紀較大的族群對線上服務的評價偏中等,如果能在簡化介面應能提升這族群對線上服務的滿意度