

第一題：

1. 程式碼：

```
setwd("C:/Users/周佳萱/Desktop/商業分析/HW3")
```

```
data = read.csv("hw3-ML.csv")
```

```
str(data)
```

```
##先把 X 和 id 這兩個無關變數拿掉
```

```
data = data[,3:25]
```

```
##dummy variable
```

```
data$Gender = ifelse(data$Gender=="Male",1,0)
```

```
data$Customer.Type = ifelse(data$Customer.Type=="Loyal Customer",1,0)
```

```
data$Type.of.Travel = ifelse(data$Type.of.Travel=="Personal Travel",1,0)
```

```
data$Class = ifelse(data$Class=="Eco Plus",2,ifelse(data$Class=="Eco",1,0))
```

```
data$satisfaction = as.factor(ifelse(data$satisfaction=="satisfied",1,0))
```

```
##滿意度滿意為 1，不滿意為 0
```

```
#install.packages("FSelectorRcpp")
```

```
library(FSelectorRcpp)
```

```
info = information_gain(x = data[,1:22],y = data$satisfaction)
```

```
barplot(info$importance,names = info$attributes,las = 2, cex.names=0.7)
```

```
##由圖我們可以發現 Type.of.Travel/Class/Inflight.wifi.service/Online.boarding/
```

```
##Inflight.entertainment/Seat.comfort 的 information_gain 較高，所以選擇這幾個變數
```

```
##1
```

```
airline = data.frame(Type.of.Travel = data$Type.of.Travel, Class = data$Class
```

```
                , Inflight.wifi.service = data$Inflight.wifi.service
```

```
                , Online.boarding = data$Online.boarding
```

```
                , Inflight.entertainment = data$Inflight.entertainment
```

```
                , Seat.comfort = data$Seat.comfort
```

```
                , satisfaction = data$satisfaction)
```

```
head(airline)
```

```
##使用 random forest 預測
```

```
library(randomForest)
```

```
rf = randomForest(satisfaction~., data = airline, importance=TRUE, ntree = 200)
```

```
#prediction
```

```
pred=predict(rf, newdata = airline[1:6])
```

```
head(pred)
```

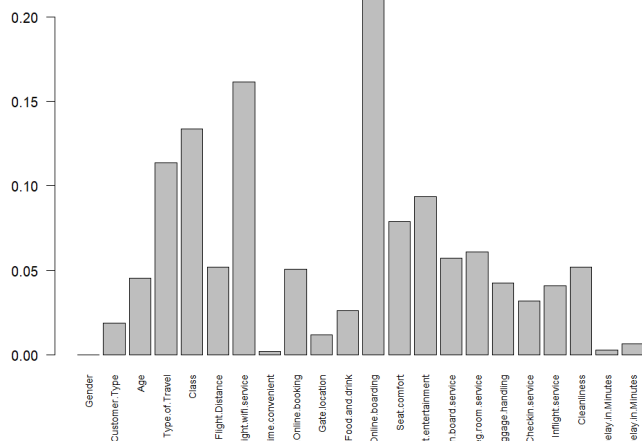
##選擇較重要的變數

importance(rf)

varImpPlot(rf)

#inflight.wifi.service、Online boarding、Type of Travel 的值較高

2. 結果敘述：



圖一、變數的 information gain 圖

首先，我們先將一些類別型變數進行 dummy variables，接著，使用 information\_gain 的套件，選出變數中 information gain 較高前六個來做分析，如圖一。我最後選擇 Type.of.Travel、Class、Inflight.wifi.service、Online.boarding、Inflight.entertainment、Seat.comfort 共六個變數來代表資料。接下來，使用 random forest 模型來預測滿意度，前六筆為：

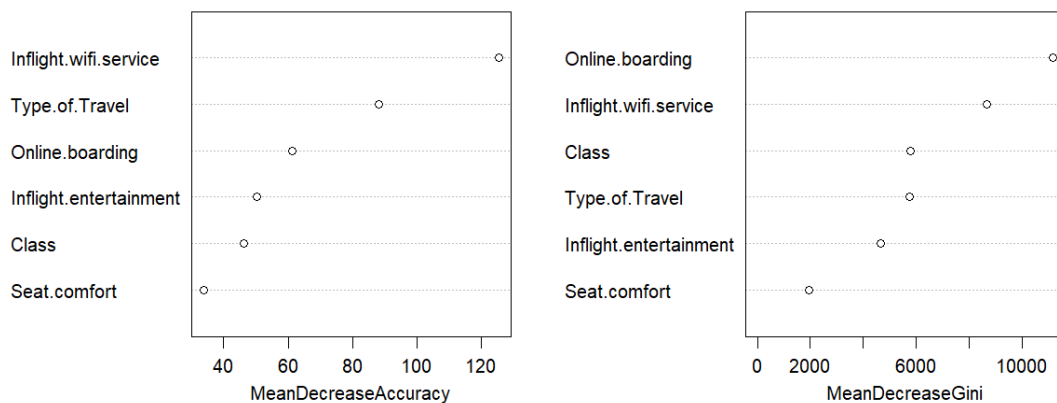
```
>head(pred)
```

```
1 2 3 4 5 6
```

```
0 0 1 0 1 0
```

```
Levels: 0 1
```

rf



圖二、variable importance in random forest 圖

由圖二，我們可以發現 inflight.wifi.service、Online boarding、Type of Travel 在隨機森林的模型中，是重要的影響預測的變數。

第二題：

1. 程式碼：

```
airline_data = data.frame(Flight.Distance = data$Flight.Distance, Age = data$Age
                          , Type.of.Travel = data$Type.of.Travel
                          , Class = data$Class, satisfaction = data$satisfaction
                          , Inflight.wifi.service = data$Inflight.wifi.service
                          , Online.boarding = data$Online.boarding
                          , Inflight.entertainment = data$Inflight.entertainment
                          , Seat.comfort = data$Seat.comfort)
```

#標準化

```
airline_data$Age = (airline_data$Age - min(airline_data$Age)) / (max(airline_data$Age))
airline_data$Flight.Distance = (airline_data$Flight.Distance - min(airline_data$Flight.Distance)) /
(max(airline_data$Flight.Distance))
```

#隨機抽取 10000 筆

```
set.seed(1)
n = sample(1:nrow(airline_data), 10000, replace=FALSE)
airline = airline_data[n,1:9]
```

#選出最適合的群數(決定選 3 組)

```
fviz_nbclust(airline,
             FUNcluster = kmeans,# K-Means
             method = "wss",      # total within sum of square
             k.max = 20           # max number of clusters to consider
             ) +
  labs(title="Elbow Method for K-Means")+
  geom_vline(xintercept = 3, linetype = 2)
```

#Kmeans(取年齡和飛行距離兩個變數)

```
km = kmeans(airline[1:2], centers=3, nstart=25)
km
```

```
table(km$cluster)
cc = km$cluster
data = cbind(airline,cc)
```

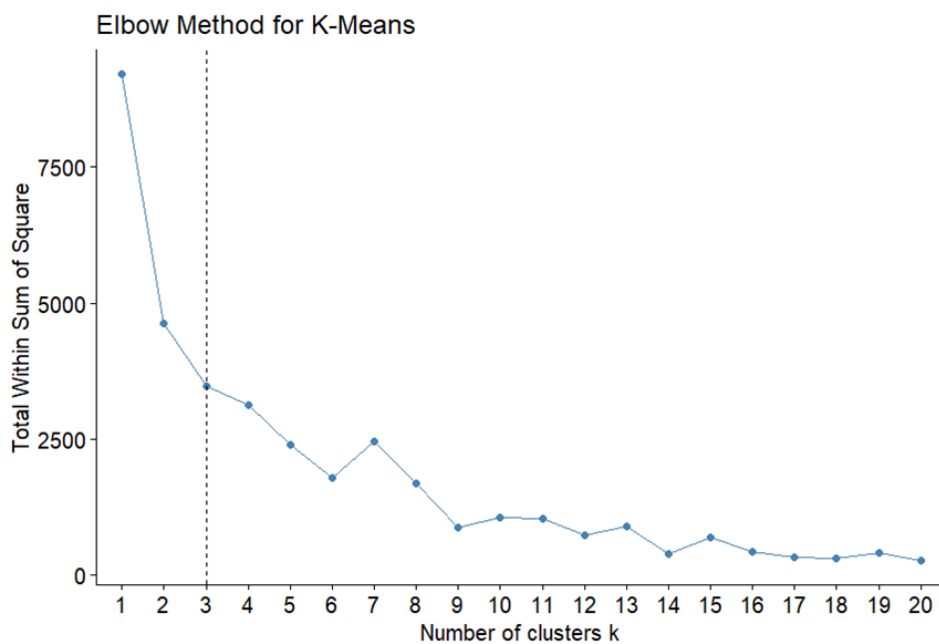
```
ggplot(data, aes(x=as.factor(cc), y=Age)) +
  geom_boxplot()
```

```

ggplot(data, aes(x=as.factor(cc), y=Flight.Distance)) +
  geom_boxplot()
ggplot( data =data) +
  geom_bar( aes( x = Type.of.Travel)) +
  facet_wrap( ~ cc)
ggplot( data =data) +
  geom_bar( aes( x = Class)) +
  facet_wrap( ~ cc)
ggplot( data =data) +
  geom_bar( aes( x = satisfaction)) +
  facet_wrap( ~ cc)
ggplot( data =data) +
  geom_bar( aes( x = Inflight.wifi.service)) +
  facet_wrap( ~ cc)
ggplot( data =data) +
  geom_bar( aes( x = Online.boarding)) +
  facet_wrap( ~ cc)
ggplot( data =data) +
  geom_bar( aes( x = Inflight.entertainment)) +
  facet_wrap( ~ cc)
ggplot( data =data) +
  geom_bar( aes( x = Seat.comfort)) +
  facet_wrap( ~ cc)

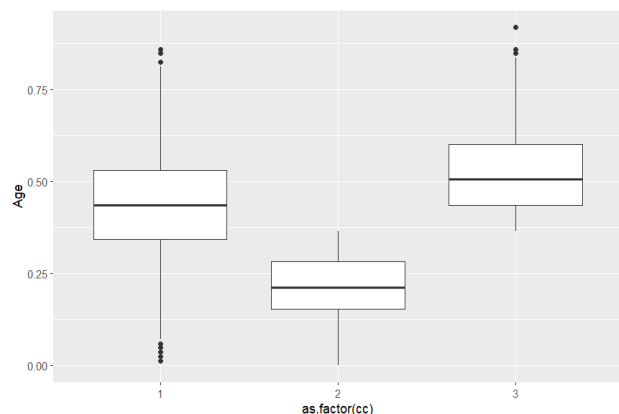
```

## 2. 結果敘述：

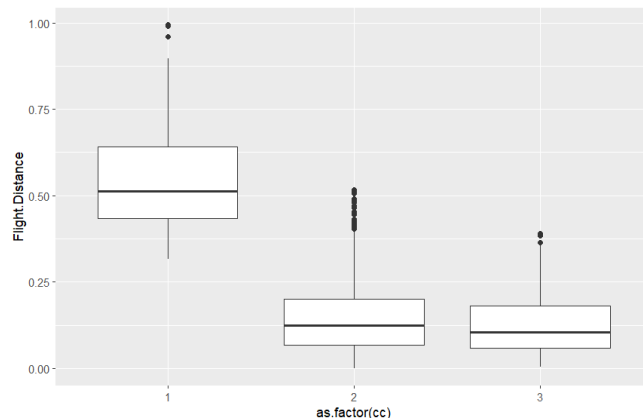


圖三、航空業客戶的 Elbow 圖

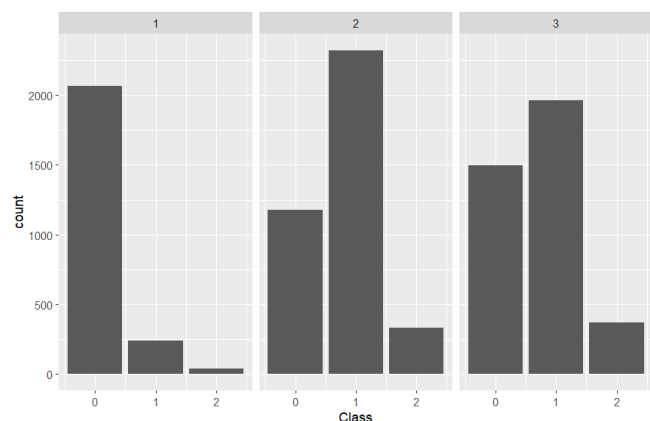
因為電腦記憶體不足，所以我選擇隨機抽取 10000 筆客戶資料，變數的部分我選擇跟客戶有關的 Age、Flight.Distance、Class、Type.of.Travel，以及跟滿意度較有關 inflight.wifi.service、inflight.entertainment、Online.boarding、Seat.Comfort，接著，我使用了 K-means 集群分析，以 Age 和 Flight.Distance 這兩個數值型變數進行分析。根據圖三，Total Within Sum of Square 在分群數 3 有顯著下降，因此，選擇 3 作為分群數。



圖四、各群航空業客戶年齡的盒鬚圖

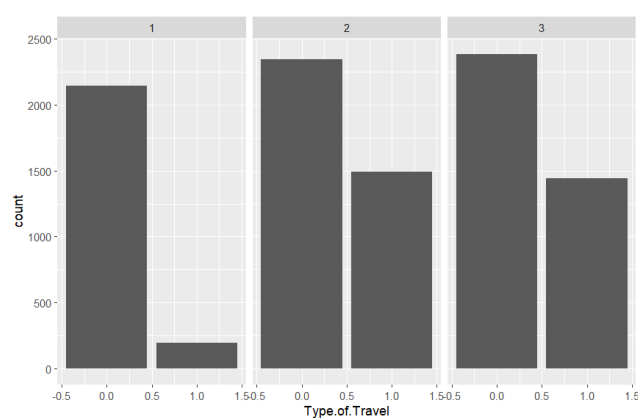


圖五、各群航空業客戶飛行距離的盒鬚圖



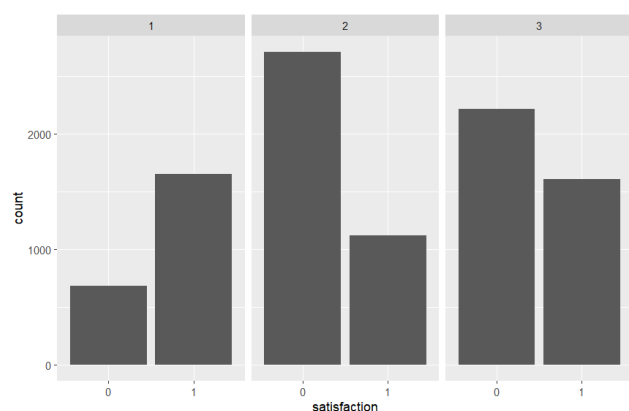
圖六、各群航空業客戶的艙位長條圖

\*0:Business,1: Eco, 2:Eco Plus



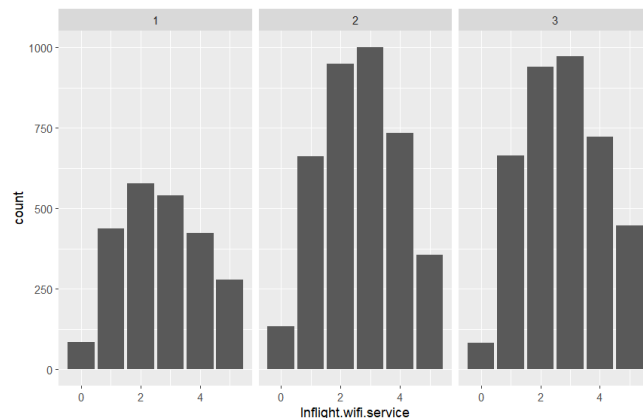
圖七、各群航空業客戶旅行目的長條圖

\*0:Business Travel、1:Personal Travel

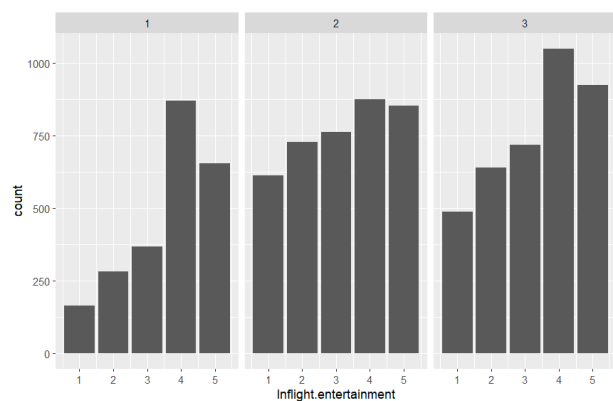


圖八、各群航空業客戶滿意度長條圖

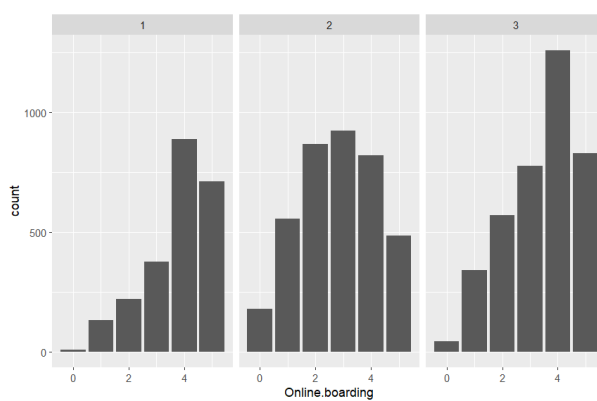
\*0:不滿意，1:滿意



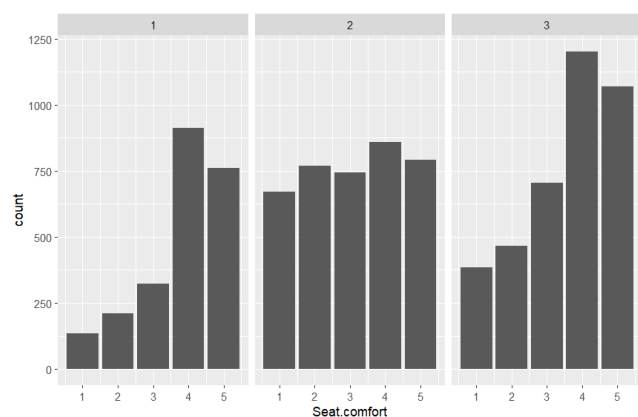
圖九、各群航空業客戶無線網路滿意度長條圖



圖十、各群航空業客戶飛機上娛樂度滿意度長條圖



圖十一、各群航空業客戶網路辦理登機滿意度長條圖



圖十二、各群航空業客戶座位舒適度滿意度長條圖

根據資料將客戶分成三群，由圖四到圖十二整理各群客戶的特色，並發想策略如下：

族群類別	客戶族群	族群特色	策略
1	長程飛行的中年客戶	<ol style="list-style-type: none"> <li>1. 多坐商務艙，搭飛機多為國外工作。</li> <li>2. 客戶對飛行體驗的滿意度高，其中對飛機上娛樂程度、辦理登機滿意度、座位舒適度滿意度較高，對飛機無線網路滿意度較低。</li> </ol>	<p>針對經常出差的族群，可以推出商務艙會員制，並且提供以下策略：</p> <ol style="list-style-type: none"> <li>1. 商務艙網路套餐：提供更良好的連線品質，且業者也要加強商務艙的無線網路設備。</li> <li>2. 商務艙享樂方案：提供客戶個人化的電影選擇，並且可以許願想要在飛機上看的電影。推出按摩椅、蒸氣眼罩、護頸枕等用品，提升客戶搭機的舒適度。</li> </ol>
2	短程飛行年輕客戶	<ol style="list-style-type: none"> <li>1. 多坐經濟艙，搭飛機可能為工作或出國。</li> <li>2. 客戶對飛行體驗的滿意度較低，其中對飛機上娛樂程度、飛機無線網路滿意度、辦理登機滿意度、座位舒適度較低。</li> </ol>	<p>年輕出國旅客可能較其他族群在意搭飛機的過程體驗，可以提供以下策略：</p> <ol style="list-style-type: none"> <li>1. 更新飛機娛樂系統，加入近幾年的遊戲、電影和音樂，提供年輕人更多娛樂選擇。</li> <li>2. 多設置快速通道、自行報到系統，提升年輕族群的登機效率。</li> </ol>

			3. 提供年輕族群里程累積活動，累積一定里程數可以用較便宜的價格升級商務艙。
3	短程飛行中老年客戶	<ol style="list-style-type: none"> <li>1. 多坐商務艙或經濟艙，可能為工作或出國。</li> <li>2. 客戶對飛行體驗的滿意度較為中立，其中對飛機上娛樂程度、辦理登機滿意度較高，飛機無線網路滿意度、座位舒適度較低。</li> </ol>	<p>中老年短程客戶對飛行體驗的滿意較為中立，可以提供以下策略：</p> <ol style="list-style-type: none"> <li>1. 座椅提供腿部調整，腰部提供按摩功能，提升搭機的舒適度。</li> <li>2. 提供大型螢幕，增加中老年人有興趣的歌謠和電影，提升老年人娛樂體驗。</li> <li>3. 提供長者服務，訓練機場人員協助長者登機前的報到流程，確保中老年在旅程中感受貼心的服務。</li> </ol>