作業一：修改鐵達尼範例程式到套用點餐判斷問題，請依講義資料(附件為train.csv檔)作分析。

範例程式

範例程式 請依你的環境做修改。train.csv在附件

import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
%matplotlib inline

from sklearn import preprocessing, tree  
#from sklearn.cross\_validation import train\_test\_split #舊版  
from sklearn.model\_selection import train\_test\_split  
from sklearn.tree import DecisionTreeClassifier, plot\_tree  
clf = tree.DecisionTreeClassifier(random\_state=0)

titanic = pd.read\_csv("train.csv")  
#Age中有NaN資料  
age\_median = np.nanmedian(titanic["Age"])   
#計算age中位數

new\_age=np.where(titanic["Age"].isnull(), age\_median, titanic["Age"])  
 #若空以中位數取敗  
titanic["Age"]=new\_age  
#PClass欄位為無\文字轉數字  
label\_encoder = preprocessing.LabelEncoder()  
encoded\_class = label\_encoder.fit\_transform(titanic["Pclass"])   
#姜滄等轉繩數字 1st, 2nd, ...  
titanic["Sex"].replace(['female','male'],[0,1],inplace=True)   
#將female male 轉成 0, 1  
X= pd.DataFrame([titanic["Sex"], encoded\_class]).T  
 #Sex為string   
X.columns=["Sex", "Pclass"]  
#X= pd.DataFrame([encoded\_class,  titanic["Age"]]).T  
y = titanic["Survived"]

Xtrain, XTest, yTrain, yTest = \  
train\_test\_split(X, y, test\_size=0.25, random\_state=1)  
dtree =tree.DecisionTreeClassifier()  
dtree.fit(Xtrain, yTrain)  
print("準確率 :", dtree.score(XTest, yTest))  
preds= dtree.predict\_proba(X=XTest)  
print(pd.crosstab(preds[:,0], columns=[X["Pclass"],XTest["Sex"]]))

preds= dtree.predict\_proba(X=XTest)  
print(pd.crosstab(preds[:,0], columns=[X["Pclass"],XTest["Sex"]]))

plt.figure()  
plot\_tree(clf, filled=True)  
plt.show()