作業二：修改鐵達尼範例程式到套用點餐判斷問題，請依講義資料，自行建立 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()
```