學號：0653440

載入的套件：

```
import pandas as pd
import graphviz
from sklearn import tree
from sklearn.model_selection import train_test_split
```

1 將資料讀取進來(可用 pandas 套件)

```
df = pd.read_csv("character-deaths.csv")#讀取資料
```

將 csv 資料讀取進來存在 df。

2 將資料處理(請用程式碼處理)

2-1 把空值以 0 替代

```
df=df.fillna(0)#把空值以0替代
```

2-2Death Year , Book of Death , Death Chapter 三者取一個，將有數值的轉成 1

```
df['Death Year'][df['Death Year'] != 0] = 1#Death Year將有數值的轉成1
```

2-3 將 Allegiances 底下的家族轉成 dummy 的特徵(底下有幾種分類就會變成幾個特徵，值是 0 或 1，本來的資料集就會再增加好幾十種特徵)

```
df = pd.get_dummies(df, columns=['Allegiances'])#將Allegiances底下的家族轉成dummy的特徵
```

3 使用 python 中 scikit-learn 上的 decision tree(前 75%是訓練集、後 25%當測試集,可以先試著將網頁範例(iris)跑出來在使用這次作業的資料集)

```
X = df.drop('Death Year', axis=1)
y = df['Death Year']
X_train, X_test, y_train, y_test = train_test_split(X, y,train_size=0.75,test_size=0.25)#前75%是訓練集、後25%當測試集
model = tree.DecisionTreeClassifier()
model = model.fit(X_train, y_train)
```

4 計算 Precision Rate , Recall Rate , Accuracy

```
y_predict = model.predict(X_test)
from sklearn.metrics import accuracy_score
from sklearn.metrics import recall_score
from sklearn.metrics import precision_score
print('Precision Rate =',precision_score(y_test, y_predict))
print('Recall Rate =',recall_score(y_test, y_predict))
print('accuracy =',accuracy_score(y_test, y_predict))
```

5 產出決策樹的圖

```
dot_data = tree.export_graphviz(model, out_file=None,max_depth=6)#限制樹的深度，以免結果無法顯示
graph = graphviz.Source(dot_data)
graph.render("df", view=True)
```