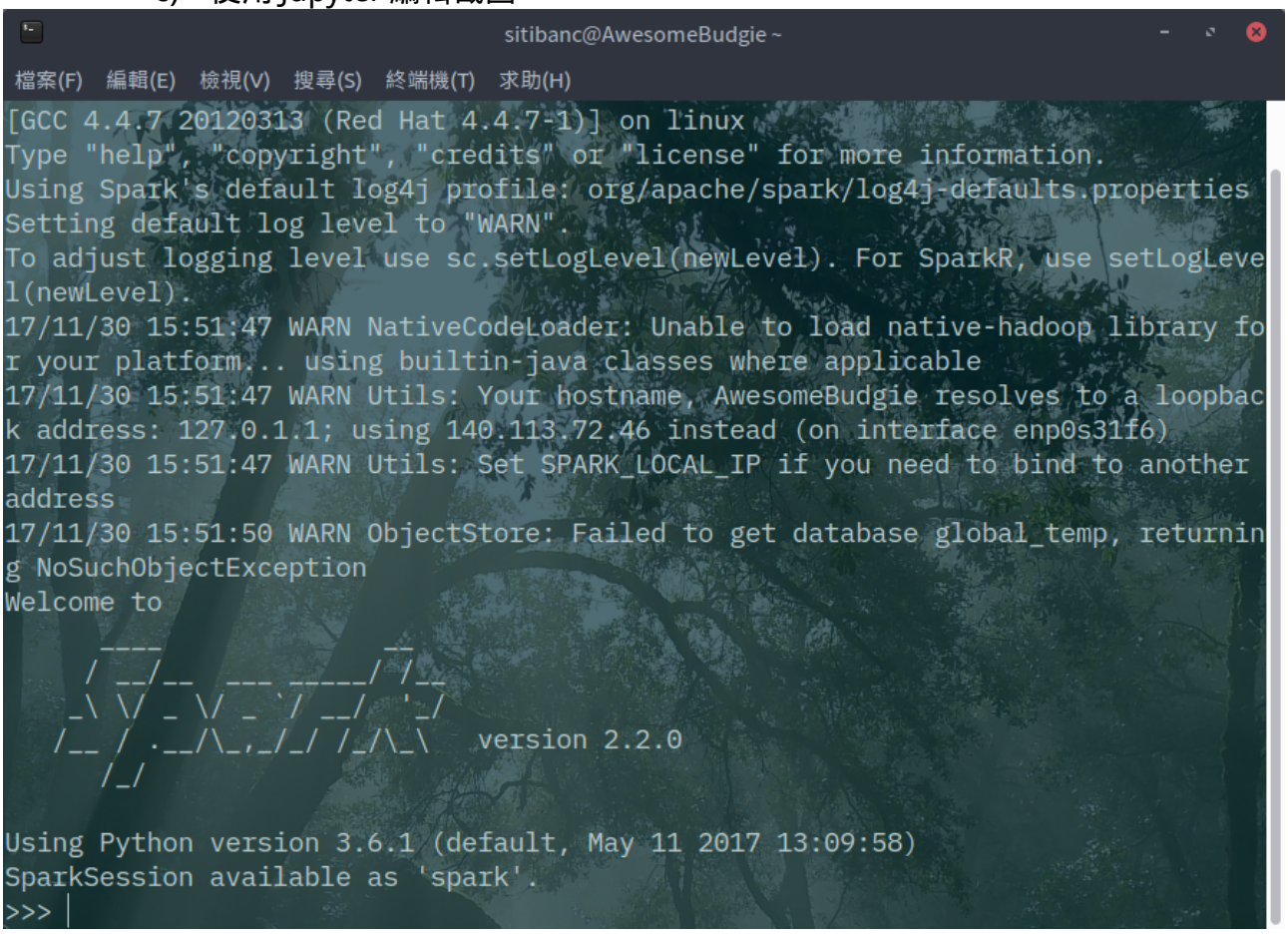
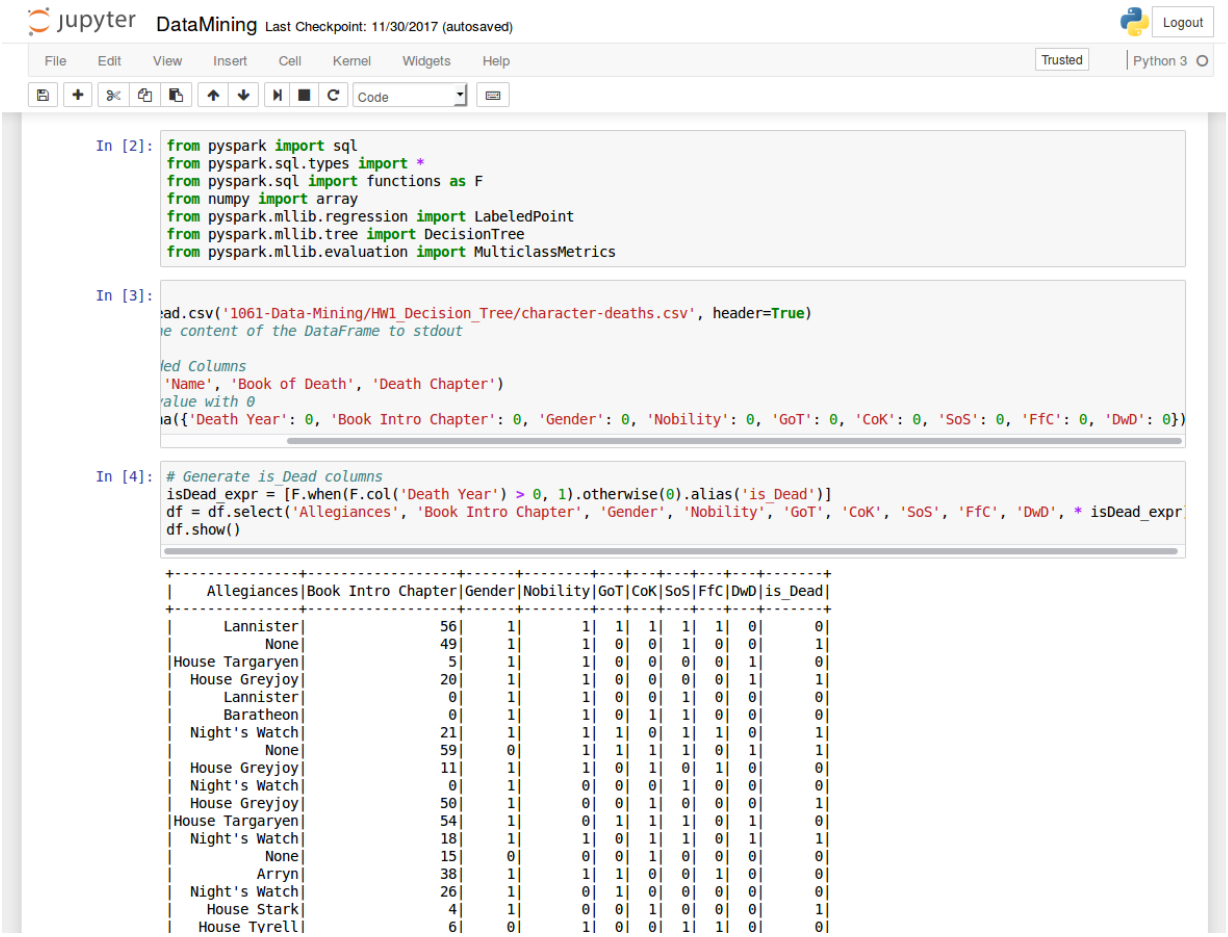
Spark 安裝

1. VirtualBox 安裝
   a) 無（本來就是 Linux 環境）
2. Spark 與 Hadoop 安裝（http://simp.ly/p/G5Lpmx）
   a) 上面的連結為安裝過程中使用到的指令紀錄及說明
   b) pyspark 執行畫面
   c) 使用 Jupyter 編輯截圖



（編輯 bashrc，加入以下內容，可在 terminal 輸入 pyspark 時自動開啟 jupyter）

```
export PYSPARK_DRIVER_PYTHON="jupyter"
export PYSPARK_DRIVER_PYTHON_OPTS="notebook"
```



程式結果

Precision: 52.083333333333336 %
Recall: 64.1025641025641 %
Accuracy: 68.51063829787233 %

討論

與作業一相同的前處理（使用 pyspark），但作業一和本次作業在切割訓練及測試資料集時都是使用 RandomSplit 的方式，因此兩次的結果略有不同。