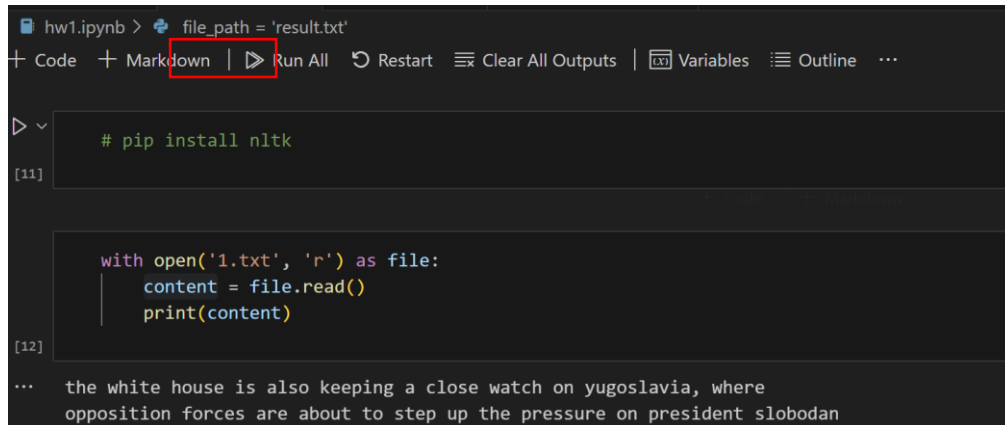


Text mining PA1

B11705058 資管三 廖振翔

1. 執行環境: Jupyter Notebook
2. 程式語言: Python 3.12.6
3. 執行方式: 執行前先 install nltk 套件，使用“pip install nltk”，接著直接按 Run All 執行即可。



```
hw1.ipynb > file_path = 'result.txt'
+ Code + Markdown | Run All Restart Clear All Outputs Variables Outline ...

[11] # pip install nltk

[12] with open('1.txt', 'r') as file:
      content = file.read()
      print(content)

... the white house is also keeping a close watch on yugoslavia, where
    opposition forces are about to step up the pressure on president slobodan
```

4. 作業處理邏輯說明: 照 ppt 上作業說明所說，先讀入檔案，把標點符號去掉並 Tokenization，全部轉換成小寫後，import nltk 使用 Porter's algorithm 來 Stemming，再從網路上找 stop word list 把 stop word 移除，接著就可以輸出檔案為“result.txt”。