

國立聯合大學教育部高等教育深耕計畫

活動成果集錦

日期：111 年 09 月 24 日

填表人：胡天鐘

執行單位	經營管理學系
活動名稱	網路評語與情資分析(二)
所對應之量化 指標名稱	111 年度協助執行深耕子計畫 4-1：推動適性化培訓課程-辦理在地產業人才適性培訓活動
主講/演出	吳光耀教授
活動時間	自 111 年 09 月 24 日 09 時 00 分 起 至 111 年 09 月 24 日 12 時 00 分 止
活動地點	國立聯合大學二坪山校區立德樓 C1-103 電腦教室(苗栗縣苗栗市恭敬里聯大一號)
參與人數	教師： <u>2</u> 人、學生： <u>8</u> 人、行政人員： <u>0</u> 人、校外： <u>31</u> 人
活動目的	在地產業人才適性培訓活動-網路評語與情資分析(二)
活動內容	<ol style="list-style-type: none"> 1. Gmap_1 評論爬文 2. Gmap_2 資料分析 3. Gmap_3 評語情緒 4. Q&A


```

    天前") if i != -1: return(int(P[:i])*1) i = P.find(" 小時前") if i != -1:
return(int(P[:i])/24) author = [] grade = [] posttime = [] postdays = [] comment =
[] commsize = [] for p in range(98): sleep_time=random.uniform(0.8,3.9) print(f"
【 {p*10:3d}】 暫停{sleep_time:.2f}秒") time.sleep(sleep_time) text =
requests.get(url1+str(p*10)+url2).text pretext = ')}\' text = text.replace(pretext,"
soup = json.loads(text) conlist = soup[2] if conlist==None: break for Con in
conlist: author.append(str(Con[0][1])) i d( (C [1])) posttime.append(str(Con[1]))
postdays.append(trandays(str(Con[1]))) posttext =
str(Con[3]).replace('\n',"").replace(' ','').replace('(由 Google 提供翻譯)',").re
postsize = 0 if Con[3]==None else len(posttext) comment.append(posttext)
commsize.append(postsize) grade.append(str(Con[4])) print("%12s %2s %4d
%30s " % (str(Con[1]), str(Con[4]), postsize, str(Con[0][1]))) #爬文清洗原始碼
https://ithelp.ithome.com.tw/questions/10204654

```

2. Gmap_2 資料分析.ipynb

目的：分析商家的評論

- 欄位：評論者、時間、評語、評等、天數、字數
- 檢視內文是否有疑義
- 檢查字數的異常值與數值分佈
- 擬定計量方法

如何使用本程式

- 投入：上傳 1 個檔 > Gmap 評論爬文.csv 【注意：編碼應存成 UTF-8】
- 產出：螢幕顯示圖表，自行複製貼到.pptx 簡報檔

```

# 載入套件：數值處理，資料表，繪圖，線性回歸 import numpy as np import
matplotlib.pyplot as plt import pandas as pd import seaborn as sns from
sklearn.linear_model import LinearRegression from scipy import stats from
scipy.stats import norm

```

```

# Colab 進行 matplotlib 繪圖時顯示繁體中文

```

```

# 下載台北思源黑體並命名 taipei_sans_tc_beta.ttf，移至指定路徑 !wget -O
TaipeiSansTCBeta-Regular.ttf
https://drive.google.com/uc?id=1eGAsTN1HBpJAkeVM57_C7ccp7hbgSz3_&ex
po import matplotlib as mpl import matplotlib.pyplot as plt from
matplotlib.font_manager import fontManager # 改 style 要在改 font 之前

```

```

# plt.style.use('seaborn') fontManager.addfont('TaipeiSansTCBeta-Regular.ttf')
mpl.rc('font', family='Taipei Sans TC Beta')

```

```

# 讀取資料 並找出字數為 0 者 gmap = pd.read_csv("Gmap 評論阿蘭.csv",
index_col=0, encoding='utf-8-sig')print('共讀取', len(gmap), '筆')

```

```

    共讀取 1920 筆

```

```

# 觀看字數為 0 者之評等是否有特別現象 idx = gmap[gmap["字數
"]==0].indexprint(gmap.loc[idx,'評等'].value_counts())

```

```

Series([], Name: 評等, dtype: int64)

```

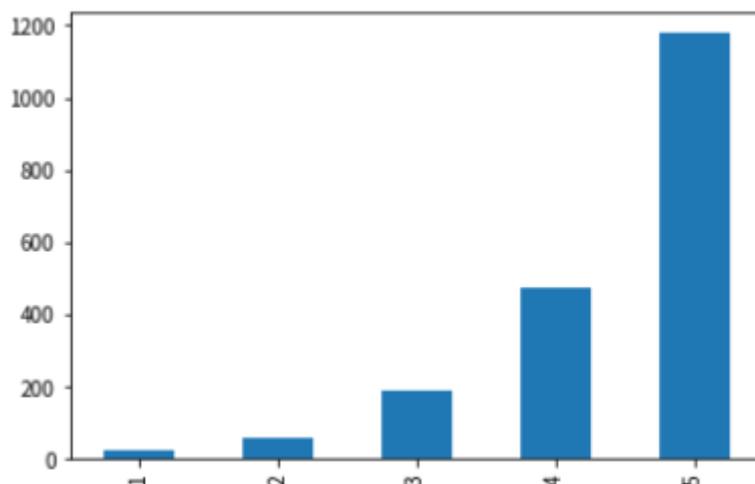
```
# 字數為 0 者似無參考價值，刪除此類訊息，並顯示前後兩筆以確認資料內容
idx = gmap[gmap["字數"]==0].index
if len(idx) > 0: gmap.drop(idx,
inplace=True)
glen = len(gmap)
display(gmap)
```

	評論者	時間	
0	小晴天	7 個月前	《阿蘭姐小吃店》苗栗市區銅板美食 😊 難得走訪苗栗必須吃在地人推薦一出火車站馬
1	幕後黑手	6 個月前	「阿蘭姊」更換了招牌，第三代接班後把這間小店重新做了整理，換上的新招牌讓人有
2	eN R	3 週前	假日過來要小排隊一下，不過等待時間不會太久，因為都是小吃的關係，翻桌率滿快的
3	OuLuu	5 個月前	苗栗人推薦啦！！！！這附近賣炒麵、水晶餃的都吃過一輪了～還是覺得阿蘭姐比較好
4	Chris Chou	3 週前	苗栗在地傳統美食，平價又好吃，水晶餃內餡調味剛剛好，不用沾醬就很夠味，難怪 /
...	
1915	Jason Lee	8 個月前	看外地人來苗栗幾乎都是來吃阿蘭姐 😊 但其實附近店家賣的幾乎都大同小異，差別在:
1916	KEVIN Hu	3 年前	很傳統的客家美食，值得

```
# 敘述統計表 gmap.describe(include='all')
```

【直方圖】評等

```
gmap['評等'].value_counts().sort_values().plot(kind='bar', color=['gray']) # 依據
次數排列(參考而已)
plt.show()
```



```
# 峰度與偏態，常態分佈的標準值 = 0
```

```
from scipy.stats import skew, kurtosis
```

```
x = gmap['評等'].value_counts().sort_index()
print("偏態 %.4f 峰度 %.4f" %
      (skew(x), kurtosis(x)))
```

```
# 是否為常態分佈：但必須 8 個以上區間才能分析：本案例不適用
```

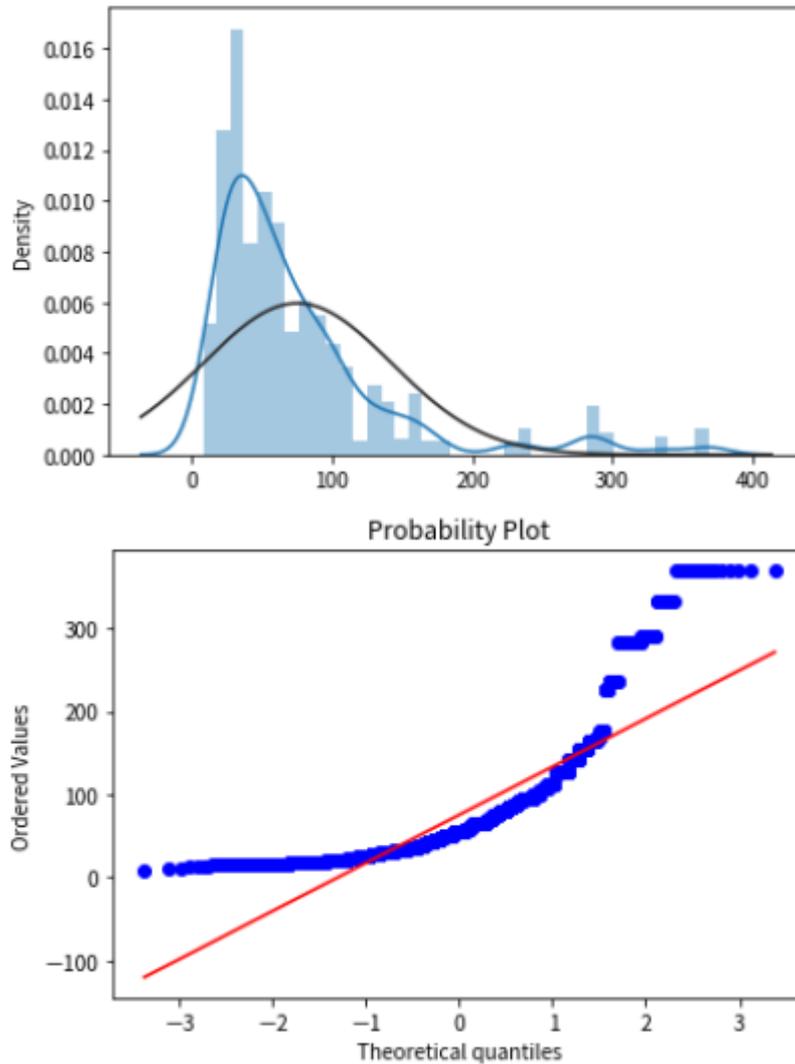
```
# from scipy.stats import normaltest
```

```
# x = gmap['評等'].value_counts().sort_index()
```

```
# print(normaltest(x))
```

偏態 1.0606 峰度 -0.4311

【直方圖，分位圖和箱型圖】字數



3. Gmap_3 評語情緒

目的：驗證語意是否顯著影響評等

- 欄位：評論者、時間、評語、評等、天數、字數
- 剖析評語的正負情緒，將之量化為語意
- 檢查語意的異常值與數值分佈
- 線性迴歸分析：模式是否成立、自變數是否顯著
- 迴歸殘差檢定

如何使用本程式

- 投入：上傳 4 個檔 > Gmap 評論爬文.csv、dict.txt.big、NTUSD_positive_unicode.txt、NTUSD_negative_unicode.txt
- 產出：螢幕顯示圖表，自行複製貼到.pptx 簡報檔

```

# 載入套件：數值處理，資料表，繪圖，線性回歸
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
from sklearn.linear_model import LinearRegression
from scipy import stats
from scipy.stats import norm
! pip install jieba
import jieba #導入中文斷詞套件(結巴)及其語庫
jieba.set_dictionary('dict.txt.big')

```

	評論者	時間	評語	評等	天數	字數	語意
0	YuChen Chiang	4 個月前	採用車牌辨識進場，停車格位非常多，但熱門時段真的一位難求，格位也不好找，有些停車格位的燈號還...	3	120	99	0.433358
1	Kevin C.	8 個月前	裝潢不錯的停車場有車牌辨識，也可以用車牌找車（機車不行）在地下室不會很悶熱手機訊號也滿格很多...	5	240	261	-0.897319
2	Google RobusTraveler	1 週前	要去市政府辦事，以前停過這裡，這次就下去了。沒想到明明電子告示板上說有許多空位，下去後一個都...	1	7	156	-0.395054
3	盧彥韶	3 週前	車格燈號完全是裝好看的新的市府停車場管理公司但跟	1	21	94	-0.438179

```

# 選擇資料
data = pd.DataFrame()
data['評等'] = gmap['評等'].values
data['字數'] = np.log(gmap['字數'].values+0.01)
data['天數'] = gmap['天數'].values
data['語意'] = gmap['語意'].values
data.tail(3)

```

	評等	字數	天數	語意
180	1	1.388791	365	0.000000
181	5	1.388791	365	0.000000
182	4	2.080691	330	0.868589

Theoretical quantiles

活動照片

111 年度高等教育深耕計畫

子計畫 4-1 促進在地產業發展、串聯地方政府及產業資源-在地產業人才適性培訓活動：「網路評語與情資分析(二)」

指導單位：教育部

主辦單位：國立聯合大學研究發展處創新育成中心、經營管理學系

計畫名稱：111 年高等教育深耕計畫

課程時間：111 年 09 月 24 日(六)上午 9 時至 12 時

課程地點：國立聯合大學二坪山校區立德樓經管系 CI-103 教室
(苗栗縣苗栗市恭敬堂聯大一號)

主持人：國立聯合大學經營管理學系 胡天鐘老師

上課教師：國立聯合大學經營管理學系 吳光耀老師

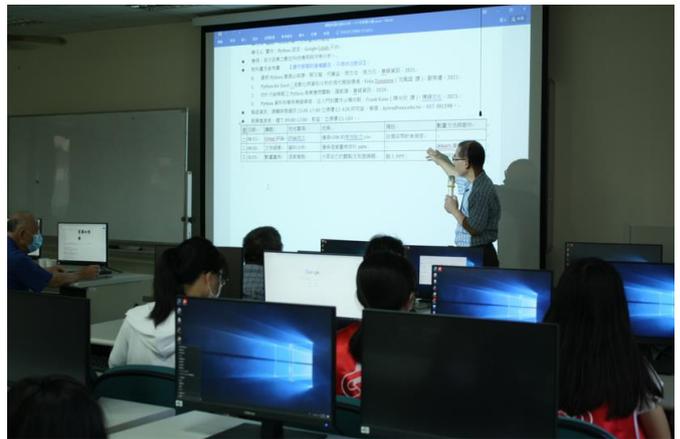
課程規劃：

時間	課程內容
09:00—09:10	主持人致詞
09:10—10:00	文字語意
10:10—11:00	資料分析與網路評論 (文字資料的處理)
11:10—12:00	獲得個案圖表存於 pptx



宣傳海報樣式

宣傳海報張貼



網路評語與情資分析(二)

網路評語與情資分析(二)



網路評語與情資分析(二)

網路評語與情資分析(二)