

# << 社群媒體分析 >>

## 第一次讀書會報告\_第十四組

### 一、分析主題：

針對民眾在日本旅行中對日本餐廳的評價分析

### 二、組員名單：

馮蕙芳\_N124080003

高健芝\_N124080006

廖胤翔\_M134610022

Deepan\_M134610032

孫郁琪\_M136020020

胡 賀\_J136020007

### 三、分析工具：

- 中山大學工作流程平台 Tarflow
- 工作流程名稱：報告1

### 四、動機

根據觀光局統計數據顯示，日本是台灣人最常選擇出國旅遊的國家之一。而旅行時為了確保能享受到良好品質的餐飲體驗，選擇適合自己的餐廳，通常會在行前先搜尋餐廳評價以避免浪費時間和金錢，也能克服語言與文化障礙，並防止踩雷。

此外，社交媒體與旅遊平台的大量興起，也讓人們更傾向參考他人經驗，與朋友或社群分享餐廳。基於上述因素，故引發本小組想了解餐廳評價在社群網路上的普遍情緒。

## 五、資料分析

資料範圍、來源、關鍵字設定：日旅-「餐廳」，時間段2024/01/01-2024/12/31

PTT 爬蟲 (6)

參數設定

homemaker(家管)  
home\_sale(房屋)  
Hsinchu(新竹)  
hypermall(賣場)  
Insurance(保險)  
iOS(iOS)  
Japan\_Travel(日旅)

任務結果

餐廳

排除關鍵字 ⓘ

以換行區隔，e.g.  
壽山動物園  
猴子  
...

搜尋起始日期

2024/01/01

搜尋結束日期

2024/12/31

儲存更改

1. PTT 爬蟲任務結果：

PTT 爬蟲 (6)

參數設定

任務結果

統計資訊

10 欄位數

1286 資料筆數

任務結果

Show 10 entries

Search:

system_id	artUrl	artTitle	artDate	artPoster	artCategory	artContent
-----------	--------	----------	---------	-----------	-------------	------------

全螢幕瀏覽

點我下載完整CSV資料

點我下載完整Rdata

點我下載完整json資料

## 2. 初步資料清洗：

(1) 通過觀察結果發現PTT的文章用一個換行符號表示“.”，兩個換行符號表示“。”，以及去除不需要的英文字母和數字，如http, com, jpeg, 00, imgur。

### ≡ 替換字串 (9)

參數設定

Input - 6

任務結果

選擇處理欄位 \*

artContent

替換字串設定 ⓘ

```
\n\n>>。  
\n>>,  
Sent from w\|+>>  
http>>  
com>>
```

選擇替換規則檔案 ⓘ

-----請選擇-----

儲存更改

## (2) 文章斷句

### ≡ 文章斷句 (12)

參數設定

Input - 9

任務結果

選擇處理欄位 \*

result

斷句符號

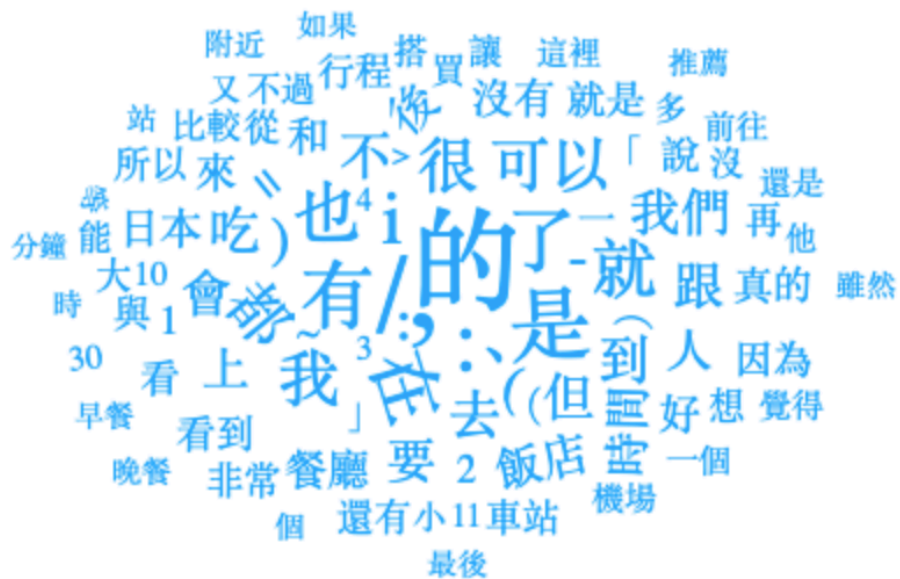
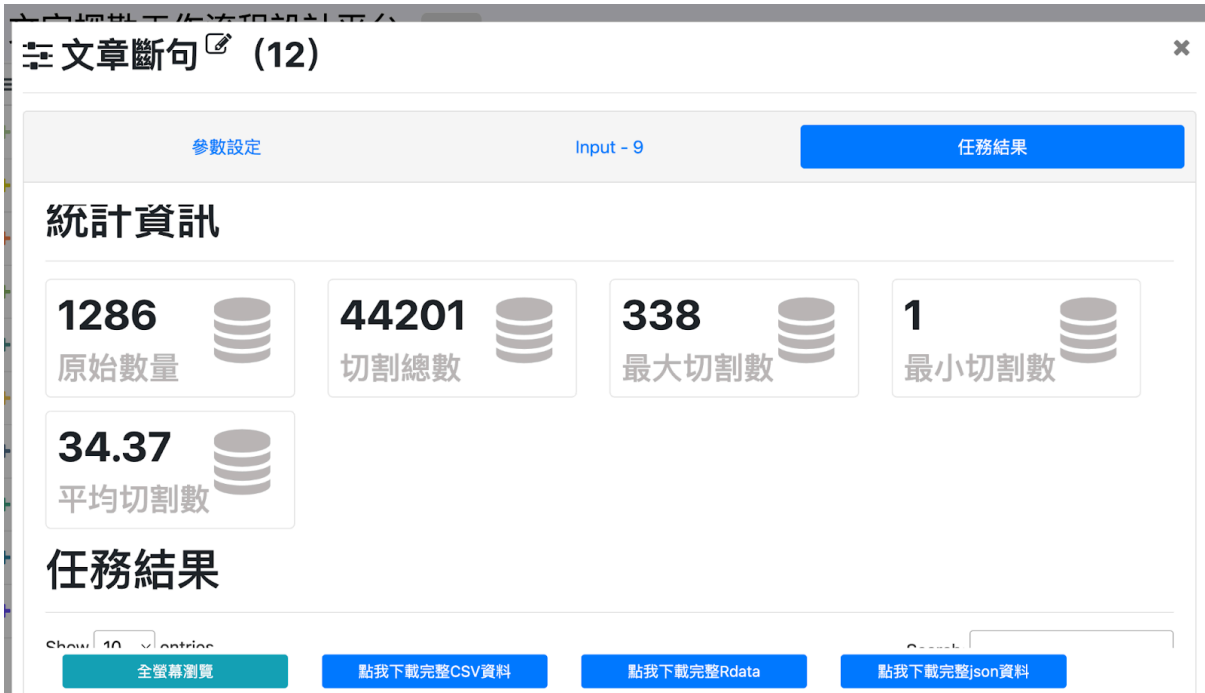
```
\n
```

使用默認符號進行斷句 \* ⓘ

是

儲存更改

(3) 中文斷詞 (ppt)



### 3. 詞性篩選

希望獲得台灣人在日旅中對餐廳的評價，因此選擇用形容詞作為篩選條件。

詞性篩選 (25)

參數設定

Input - 19

任務結果

語言 \*

Chinese

選擇保留詞性 \*

Noun  
Proper Noun  
Verb  
Adjective  
Adverb

儲存更改

### 4. 停止詞

根據形容詞篩選後的結果，自訂停止詞詞典，如下：直接、重要、主要、最後、兩張、點了、盛岡、原來、分鐘、一種。

清除停用詞 (19)

參數設定

Input - 15

任務結果

統計資訊

855372

清除總數

任務結果

Show 10 entries Search:

system\_id

sentence\_idx

result

1

1

[天婦羅, 天寅, 天寅, 用餐, 時間, 地址, 東京都, 渋谷, 渋谷, 渋谷, 交通, 渋谷, 營業, 時間,

全螢幕瀏覽

點我下載完整CSV資料

點我下載完整Rdata

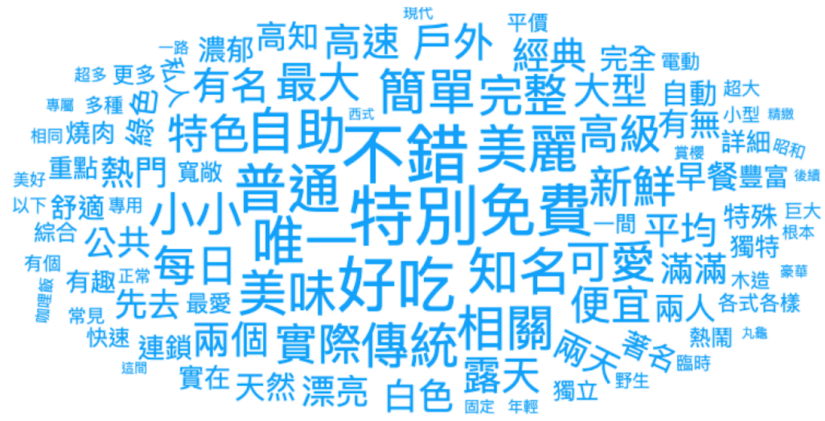
點我下載完整json資料

### 三 詞頻計算 (27)

## 參數設定

Input - 25

## 任務結果



全螢幕瀏覽

[點我下載完整CSV資料](#)

[點我下載完整Rdata](#)

[點我下載完整json資料](#)

5.我們選擇< Lexicon Based 情緒分析>, 是一種將文字轉換為情緒分數的方法，用來評估文本的情感是正面或負面。

為了能提高文本中隱含情感，能對詞彙分析更加準確，我們自己定義了一些正負面詞彙，而定義的字詞中我們輸入了一些常見的網路用語、顏文字、以及M世代及Z世代的用語。

日旅版的餐廳評論中習慣用語定義正面詞彙,包含「XD、頂級、我超、美味、好吃、入口即化、有料、很頂」,而「難吃、普通、沒味道、太鹹、太淡、油膩、乾柴、變質、冷掉」我們則定義為在餐廳的常見的負面情緒詞。

而我們產出的文字雲中顯示「推薦、好吃、喜歡、不錯」為最常出現的正向情緒詞，而「問題、普通、擔心、錯過、隨便」為出現最多次的負面情緒。

## Lexicon Based 情緒分析 (30)

參數設定

Input - 19

任務結果

語言 \*

Chinese

載入預設情緒字典 \*

是

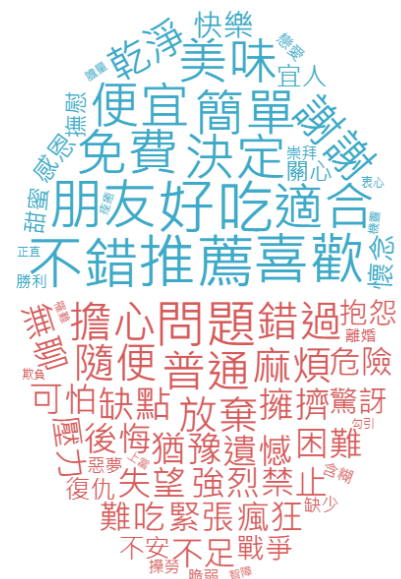
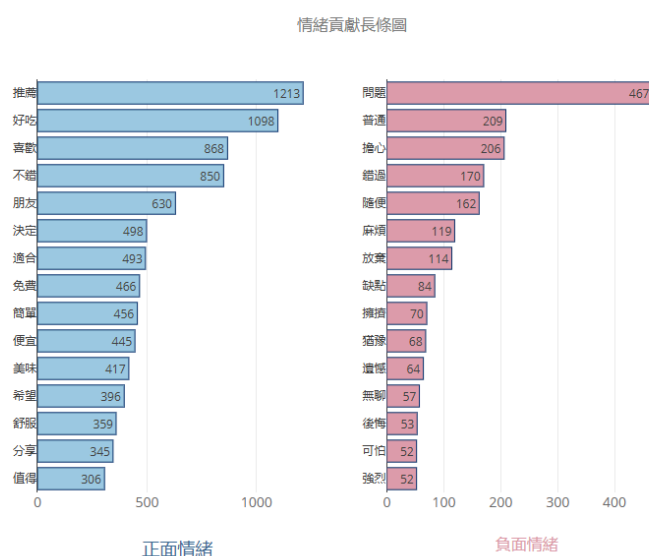
定義正面詞彙 ①

XD  
xd  
頂級  
我超  
好嫩好滑

定義負面詞彙 ①

難吃  
普通  
沒味道  
太鹹  
太淡

儲存更改



6.接著，我們以ID合併了我們的資料並劃分成不同的欄位。

### JOIN規則

新增規則

刪除規則

任務一欄位	任務二欄位
system_id	system_id
-----請選擇-----	-----請選擇-----

## 欄位篩選 (59)

參數設定

選擇要保留的欄位(按住ctrl(Windows)或command(MAC)可以複選) \*

system\_id

sentence\_idx

positive\_count

positive\_words

negative\_count

negative\_words

sentiment\_value

7.接著在新增欄位選項中我們設定了正負數量的最大值max。

## 新增欄位 (idxmax/min) (56)

參數設定

Input - 59

任務結果

匯總函數 \* ⓘ  
max

計算欄位(按住ctrl(Windows)或command(MAC)可以複選) \*  
system\_id  
positive\_count  
negative\_count  
sentiment\_value  
artDate  
artPoster

新增的欄位名稱 \*  
senti\_class



8.接著我們設定了日期，我們決定探討2024年的每個月的情緒值，而結果顯示每個月的負面情緒值大多為三位數，而正面情緒詞大多為四位數。

### 分群彙總(日期) (38)

參數設定

Input - 56

任務結果

選擇日期欄位 \*

artDate

選擇日期類型 \*

月

計算欄位(按住ctrl(Windows)或command(MAC)可以複選) \*

system\_id  
positive\_count  
negative\_count  
sentiment\_value  
artDate  
artPoster  
senti\_class

日期格式 \* ⓘ

%Y-%m-%d %H:%M:%S%z

匯總函數 \* ⓘ

sum

保留欄位(按住ctrl(Windows)或command(MAC)可以複選) \*

system\_id  
positive\_count  
negative\_count  
sentiment\_value  
artDate  
artPoster  
senti\_class

### 統計資訊

24

群組數量



### 任務結果

Show 10 entries

senti_class	artDate	sentiment_value
negative_count	2024-01	-209
negative_count	2024-02	-148
negative_count	2024-03	-204
negative_count	2024-04	-229
negative_count	2024-05	-197

分群彙總(日期) ( 38 )

參數有做更動，建議重新執行

參數設定	Input - 56	任務結果
negative_count	2024-11	-164
negative_count	2024-12	-286
positive_count	2024-01	1579
positive_count	2024-02	1168
positive_count	2024-03	1730
positive_count	2024-04	1612
positive_count	2024-05	1522
positive_count	2024-06	1786
positive_count	2024-07	1791
positive_count	2024-08	1293

Showing 11 to 20 of 24 entries

Previous123Next

接著，在資料篩選裡我們將情緒值設為小於1，因為我們只想要在這個步驟裡探討每月中立和負面的文章趨勢。

資料篩選 (62)

參數設定

Input

條件式 \*

sentiment\_value < 1

負面文章趨勢 (65)

參數有做更動，建議重新執行

參數設定

Input - 62

任務結果

選擇日期欄位 \*

artDate

選擇日期類型 \*

月

計算欄位(按住ctrl(Windows)或command(MAC)可以複選) \*

system\_id

positive\_count

negative\_count

sentiment\_value

artDate

artPoster

日期格式 \* ⓘ

%Y-%m-%d %H:%M:%S%z

匯總函數 \* ⓘ

count

保留欄位(按住ctrl(Windows)或command(MAC)可以複選) \*

system\_id

positive\_count

negative\_count

sentiment\_value

artDate

artPoster

這是負面文章趨勢的數值。

負面文章趨勢 (65)

參數有做更動，建議重新執行

參數設定

Input - 62

統計資訊

12

群組數量

任務結果

Show 10 entries

artDate	system_id
2024-01	2365
2024-02	1817
2024-03	2613
2024-04	2573
2024-05	2465

9.在欄位重構(Melt)步驟裡我們設定探討每個月的正負總數，之後我們在視覺化儀表板裡的參數設定中，選擇了< 70-分群彙總(日期) > 並把匯總函數設定為總和SUM，並且重新命名了X軸及Y軸。最後我們生出了以下圖表，圖表顯示每個月的正負總數及其變化趨勢。從圖表中我們可以發現...

## 欄位重構(Melt) (68)

參數設定

Input - 59

任務結果

要轉換的欄位(按住ctrl(Windows)或command(MAC)可以複選) \*

positive\_count

negative\_count

sentiment\_value

artDate

artPoster

要保留的欄位(按住ctrl(Windows)或command(MAC)可以複選)

positive\_count

negative\_count

sentiment\_value

artDate

artPoster

轉換的欄位新的欄位名稱 \*

melt\_senti\_count

值的欄位名稱 \*

melt\_senti\_value

## 分群彙總(日期) (70)

參數設定

Input - 68

任務結果

選擇日期欄位 \*

artDate

選擇日期類型 \*

月

計算欄位(按住ctrl(Windows)或command(MAC)可以複選) \*

system\_id

artDate

melt\_senti\_count

melt\_senti\_value

日期格式 \* ⓘ

%Y-%m-%d %H:%M:%S%z

匯總函數 \* ⓘ

sum

保留欄位(按住ctrl(Windows)或command(MAC)可以複選) \*

system\_id

artDate

melt\_senti\_count

melt\_senti\_value

70-分群彙總(日期) (70)

參數設定

Input - 68

任務

統計資訊

24

群組數量

任務結果

Show 10 entries

melt_senti_count	artDate	melt_senti_value
negative_count	2024-01	406
negative_count	2024-02	282
negative_count	2024-03	405
negative_count	2024-04	410
negative_count	2024-05	353

參數設定

圖片預覽

70-分群彙總(日期)

Group column:

melt\_senti\_count

Y 軸標籤:

artDate

Y 軸標籤: 正負數值彙總:

melt\_senti\_value

選擇函數:

sum

顯示數量:

20

圖片標題設定:

請輸入圖片標題...

X 軸標籤設定:

月份

Y 軸標籤設定:

正負總數

正負總數

positive\_count

negative\_count

Jan 2024

Mar 2024

May 2024

Jul 2024

Sep 2024

Nov 2024

月份



## 六、結論：

1. 透過長條圖表清晰顯示，每個月的正面情緒值（橘色）都遠高於負面情緒值（藍色）。這表示在 PTT 日旅版關於「餐廳」的討論中，正面評價明顯佔據主導地位。

這可能反映出台灣人對於日本旅遊餐廳的整體滿意度較高，或是正面經驗更傾向於被分享。

2. 情緒波動與月份的關聯：

雖然正面情緒整體較高，但各月份之間仍存在些許波動。例如，3月、6月、7月、10月的正面情緒值相對較高，而2月、5月、9月則相對較低。

這種波動可能與旅遊淡旺季、特殊節日、或是其他突發事件有關。例如，1月和11月可能是旅遊旺季，而2月和9月可能是旅遊淡季。

3. 負面情緒的穩定性：

相較於正面情緒的波動，負面情緒值在各月份之間相對穩定，大多維持在三位數。

負面評價的具體文字內容，例如問題、擔心、錯過、麻煩等。可能包含用餐體驗與期望不符，餐廳沒有達到預期的水準、可能擔心餐廳的價格是否合理、預約過程繁瑣等麻煩的問題。

總和來說，我們的分析顯示：

- 台灣旅客普遍對日本餐廳持正面評價，且願意分享良好體驗。
- 旅遊淡旺季可能影響餐廳評價，這是一個值得進一步探討的議題。

