

課程：社群媒體分析

第 22 組期中專案報告

組員：

N104220009 周嵩智

N104220010 吳尚蓉

N104220024 林素芬

N104220026 陳嫻雅

## A. 動機和分析目的

### ● 研究動機：

台灣作為一個美食文化豐富的地區，各地的小吃具有獨特的地理分布和口味特點。在台灣的 PTT 美食論壇上，許多網友分享和討論他們對各類小吃的喜好和見解。然而，過去研究尚未充分利用這些寶貴的數據資源，深入探討台灣小吃的地理分布和口味特點。因此，本研究以 "社群媒體分析" 課程專案報告為契機，利用 PTT 美食論壇的數據，全面分析台灣地區的小吃地理分布和口味特點，以期提供對台灣美食文化和產業的新視角和啟示。

### ● 研究目的：

- a. 描述並分析台灣地區小吃在 PTT 美食論壇上的地理分布特徵，揭示各地區小吃的特點和差異，以及可能的成因。
- b. 系統性地探討台灣地區小吃在 PTT 美食論壇上的口味特點，包括甜、辣、鹹、酸、苦等，並分析這些口味特點與地理分布的關係。
- c. 比較台灣地區小吃在 PTT 美食論壇上的評價和受歡迎程度，以了解各地區口味特點對民眾喜好的影響。
- d. 從地理分布和口味特點的角度，分析台灣特色小吃對於當地文化、歷史和社會經濟的影響及其互動關係。
- e. 提出相關建議，以促進台灣美食產業的發展和文化傳承，並為未來有關台灣地區小吃地理分布和口味特點的研究提供參考。

## B. 資料集的描述

- 來源：PTT 美食版/台中版/台南版/高雄版/八卦版

PTT 爬蟲 ( 107 )

參數設定

任務結果

選擇看板

SENIORHIGH(高中)  
Soft\_job(軟體工作)  
Steam(Steam)  
Stock(股票)  
studyteacher(實習教師)  
TaichungBun(台中)  
Tainan(台南)

搜尋關鍵字  
小吃

排除關鍵字  
以換行區隔，e.g.  
泰山動物園  
猴子  
...

搜尋起始日期  
2022/01/01

搜尋結束日期  
2023/04/16

關鍵字:小吃

資料日期:2022/01/01-2023/4/16

PTT 爬蟲 ( 107 )

參數設定

任務結果

統計資訊

10  
欄位數

4673  
資料筆數

任務結果

Search:

system\_id

artUrl

artTitle

artDate

artPoster

artCategory

artContent

artComr

[食記]壽  
義小吃小  
菜種類多

2022-01-

餐廳名稱：古早火雞肉飯  
消費時間：2021.12  
地址：

欄位數: 10

資料筆數:4673

欄位名稱	說明
artUrl	貼文 URL
artTitle	貼文標題
artDate	貼文日期
artPoster	貼文者
artCategory	貼文分類
artContent	貼文內容
artComment	貼文回應
e_ip	貼文 IP

insertedDate	寫入時間
dataSource	資料來源

## C. 資料分析過程

- 本次專題的工作流程為：Project1\_epsilon



- PTT 爬蟲：數據集來源為 PPT，使用 PPT 爬蟲工具進行資料擷取

PTT 爬蟲 (107)

參數設定 任務結果

選擇看板

- SENIORHIGH(高中)
- Soft\_Job(軟體工作)
- Steam(Steam)
- Stock(股票)
- studyteacher(實習教師)
- TaichungBun(台中)
- Tainan(台南)

搜尋關鍵字

小吃

排除關鍵字

以換行區隔，e.g.  
泰山動物園  
養子  
...

搜尋起始日期

2022/01/01

搜尋結束日期

2023/04/16

儲存更改

## ● 替換字串

替換字串 ( 2 )

參數有做更動，建議重新執行

參數設定

Input - 107

任務結果

選擇處理欄位 \*

artContent

選擇替換規則檔案 ⓘ

-----請選擇-----

替換字串設定 ⓘ

新竹市>>新竹  
 彰化縣>>彰化  
 花蓮市區>>花蓮  
 花蓮市>>花蓮  
 花蓮縣>>花蓮

儲存更改

- 地名統一
- 符號替換為空白

## ● 文章斷句

文章斷句 ( 71 )

參數有做更動，建議重新執行

參數設定

Input - 2

任務結果

選擇處理欄位 \*

result

斷句符號

- 
- ◎

使用默認符號進行斷句 \*

是

儲存更改

- 除使用默認符號斷句外，另外增加斷句符號●、◎

文章斷句 ( 71 )

參數設定

Input - 2

任務結果

4673

原始數量

40615

切割總數

543

最大切割數

1

最小切割數

8.69

平均切割數

任務結果

Show 10 entries

Search:

system_id	sentence_idx	sentence
1	1	餐廳名稱：古早火雞肉飯 消費時間：2021.12 地址：嘉義東區林森東路108號 電話：05-275-9260 營業時間：10:30-20:00 每人平均價位：100 可否刷卡...
1	2	古早火雞肉飯的位置就在嘉義東區的林森東路上，鄰近嘉義大學林森校區、嘉義舊監獄以及舊警察招待各位朋友如果剛好有到附近的話，可以來這裡吃飯
2	1	古早火雞肉飯的餐點種類真的超級多的，而且來吃飯的幾乎都是以在地人為...

介紹店舖：[知更鳥巢鮮果茶] [嘉義]

## ● 中文斷詞

中文斷詞 (3) 參數有做更動，建議重新執行

參數設定 Input - 71 任務結果

選擇處理欄位 \*  
sentence

定義詞彙 ①  
以換行符號區隔，e.g.  
詞彙 權重  
國立中山大學 1000  
西子灣 500  
...

選取字典 ①  
localfood.txt

儲存更改

a. 自行定義字典：localfood.txt，共 268 個詞彙

中文斷詞 (3)

參數設定 Input - 71 任務結果

統計資訊

1703 最大字數

1434503 總字數

0 最小字數

35 平均字數

任務結果

Show 10 entries Search:

system_id	sentence_idx	result
1	1	[餐廳名稱, :, 古早, 火雞肉飯, 消費, 時間, :, 2021.12, 地址, :, 嘉義, 東區, 林森東路, 108, 號, 電話, :, 05, -, 275, -, 9260, 營業, 時間, :, 10, :, 30, -, 20, :, 00, 每人, 平均價位, :, 100, 可否, 刷卡, :, 否, 有無, 包廂, :, 無, 推薦, 菜色, :, 火雞肉飯, -, 小菜, 好, 讚, 圖文, 版, :, 這間, 位於, 嘉義, 東區, 林森東路, 上, 的, 「, 古早, 火雞肉飯, 」, :, 是, 一間, 相當, 推薦, 的, 嘉義, 火雞肉飯, 以及, 嘉義, 在地, 字字珠, 小, 吃, 推薦, 古早, 火雞肉飯, 的, 餐點, 價格, 非常, 合理, :, 除了, 火雞肉飯, 之外, :, 還有, 雞肉飯, :, 而

## ● 清除停用詞

清除停用詞 (4)

參數設定 Input - 3 任務結果

語言 \*  
Chinese

是否清除單字元 ①  
是

清除英文字母 \*  
是

清除換行符號 \*  
是

清除html tag \*  
是

使用預設停止詞  
是

是否轉為小寫英文  
是

清除數字 \*  
是

清除特殊標點符號 \*  
是

自定義停止詞  
google  
出門  
適合  
官網

a. 限定在台灣地區，故設定清除英文字母

b. 因為只分析地區、小吃、食材、口感及感覺相關，故除去價格、環境、交通及其他無關的詞彙。

清除停用詞 ( 4 )

參數設定

Input - 3

任務結果

統計資訊

951037

清除總數

任務結果

Show 10 entries

Search:

system_id	sentence_idx	result
1	1	<p>[古早, 火雞肉飯, 嘉義, 東區, 林森東路, 火雞肉飯, 小菜, 嘉義, 東區, 林森東路, 古早, 火雞肉飯, 嘉義, 火雞肉飯, 嘉義, 在地, 老字號, 小吃, 古早, 火雞肉飯, 火雞肉飯, 爌肉飯, 咖哩飯, 超值, 便當, 滷品, 超多, 小菜, 古早, 火雞肉飯, 火雞肉, 片飯, 半熟, 荷包蛋, 黃澄澄, 蛋黃, 流出來, 口感, 超嫩, 火雞肉, 可口]</p>

## ● 詞頻計算(一)

☰ 詞頻計算 (一)  (59)

參數有做更動，建議重新執行

參數設定

Input - 4

任務結果

自定義停止詞

以換行符號區隔，e.g.  
你好  
不要  
...

最多篩選詞彙數量 

200

儲存更改

- a. 清除停用字之後，執行詞頻計算，依據詞頻清單，回頭更新停用字，反覆迭代更新至只剩所需的詞彙。

[illegible]

## ● 詞性篩選

詞性篩選 adj/adv ( 129 ) 參數有做更動，建議重新執行

參數設定      Input - 4      任務結果

語言 \*  
Chinese

選擇保留詞性 \*  
Noun  
Proper Noun  
Verb  
Adjective  
Adverb

儲存更改

### a. 篩選詞性為形容詞與副詞

詞性篩選 adj/adv ( 129 )

參數設定      Input - 4      任務結果

任務結果

Show 10 entries      Search:

system_id	sentence_idx	result
1	1	[古早]
1	2	[在地人, 古早, 轟大, 遙遠, 古早, 豬皮]
2	7	[懂一度, 俗氣, 難以, 正宗, 喜愛, 辛苦, 驚驚怕怕]
2	8	[好哥, 確實, 不假]
2	9	[而大部份, 四個, 一杯, 漂亮, 傳統, 俗調]
2	10	[特地, 遠遠, 正宗]
2	11	[在地人, 客氣, 親切, 實情]
2	12	[不想]
2	13	[一大張, 實情, 一杯遠, 果肉, 真正, 濃郁, 原來, 這杯, 果肉, 備用, 濃郁]

## ● 詞頻計算(二)

詞頻計算(二) ( 131 ) 參數有做更動，建議重新執行

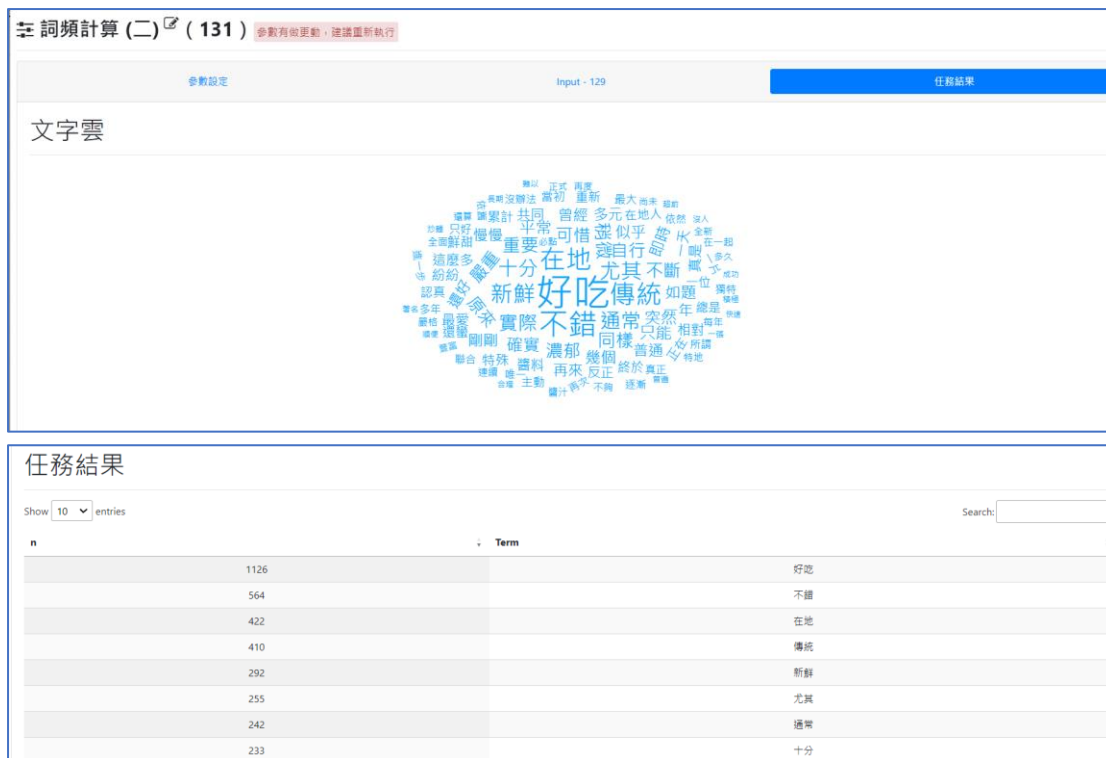
參數設定      Input - 129      任務結果

自定義停止詞  
以換行符號區隔，e.g.  
你好  
不要  
...

最多篩選詞彙數量 500

- 將經過詞性篩選為形容詞與副詞的結果，執行詞頻計算
- 下載詞頻前 500 大的形容詞與副詞
- 由人工依據該詞頻清單進行詞彙正面與負面定義，再將人為區分的定義加入 NTUSD 設定





## ● NTUSD 情緒分析

NTUSD 情緒分析 (127)

參數設定 Input - 4 任務結果

是否使用否定詞 **❗**  
否

是否使用加強詞 **❗**  
否

定義正面詞彙 **❗**  
好吃  
不錯  
在地  
傳統  
新鮮

定義負面詞彙 **❗**  
可惜  
嚴重  
還好  
普通  
不夠

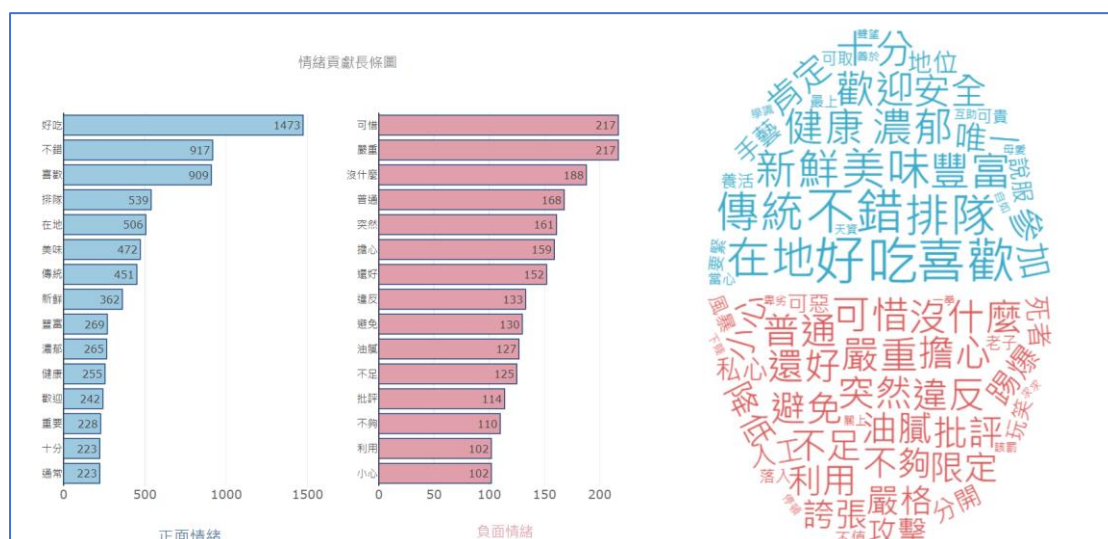
移除情緒詞 **❗**  
死亡  
刪除  
不斷  
透露  
壓力

- 依照詞頻計算(二)下載的詞頻，經過人工歸類正面與負面，將清單結果維護至此功能的正面與負面詞彙
- 定義正面詞彙：  
好吃、不錯、在地、傳統、新鮮、通常、十分、重要、濃郁、確實、再來、多元、最愛、鮮甜、唯一、依舊、真正、獨特、特地、豐富、必點、再度、成功、著名、快速、熱門、深刻、剛剛好、多樣、精緻、清淡、可口、道地、美好、特有、不貴、濃厚、熱烈、熱情、鮮美、甜甜、健康、很棒、很香、專門、安心、不膩、再訪、鮮嫩、夠味、天然、正宗、用心、排隊、不油
- 定義負面詞彙：

可惜、嚴重、還好、普通、不夠、沒人、難以、不足、太過、難吃、沒什麼、小心、膩口、踢爆、違法、人工、過度、不飽、不實

d. 移除情緒詞:

死亡、刪除、不斷、透露、壓力、減少、沒想到、辛苦、限制



## ● 合併資料

合併資料 (134)

參數設定 Input - 107 Input - 127 任務結果

JOIN規則

新增規則 刪除規則

left\_key right\_key

system\_id system\_id

-----請選擇-----

儲存更改

a. 將 NTUSD 情緒分析結果與貼文原始資料合併

## ● 分群彙總

**分群彙總 ( 137 )**

參數設定      Input - 134      任務結果

使用...欄位進行分群(按住ctrl(Windows)或command(MAC)可以複選) \*

- negative\_words
- sentiment\_value
- artUrl
- artTitle
- artDate
- artPoster
- artCategory

匯總函數 \*

- nunique
- min
- max
- first
- last
- std
- sum

計算欄位(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

a. 依照貼文日期及貼文來源(版別)進行每日情緒值分群匯總

## ● 分群彙總(日期)

**分群彙總(日期) ( 140 )**

參數設定      Input - 137      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) \*

- artDate
- artCategory
- positive\_count@sum
- negative\_count@sum
- sentiment\_value@sum

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) \*

- artDate
- artCategory
- positive\_count@sum
- negative\_count@sum
- sentiment\_value@sum

a. 將資料以月為單位分群匯總

**分群彙總(日期) ( 140 )**

參數設定      Input - 137      任務結果

統計資訊

80 群組數量

任務結果

Show 10 entries      Search:

artCategory	artDate	positive_count@sum	negative_count@sum	sentiment_value@sum
Food	2022-01	459	-154	305
Food	2022-02	488	-139	349
Food	2022-03	911	-303	608
Food	2022-04	891	-280	611
Food	2022-05	607	-222	385

## ● 詞性篩選 專有名詞

詞性篩選 專有名詞 (68) 參數有做更動，建議重新執行

參數設定

Input - 4

任務結果

語言 Chinese

選擇保留詞性

- Noun
- Proper Noun
- Verb
- Adjective
- Adverb

### a. 針對地名、小吃名、店名、路名等專有名詞進行篩選

詞性篩選 專有名詞 (68) 參數有做更動，建議重新執行

參數設定

Input - 4

任務結果

任務結果

Show 10 entries

Search:

system_id	↑ sentence_idx	↓ result
1	1	[嘉義, 林森東路, 小英, 嘉義, 林森東路, 嘉義, 嘉義, 古早, 鹽肉飯, 咖哩飯, 小英, 黃豆湯]
1	2	[嘉義, 林森東路, 嘉義, 林森, 嘉義, 小英, 湯米, 划算, 千葉別]
2	1	[嘉義]
2	2	[嘉義, 西區]
2	7	[嘉義, 台灣, 嘉義]
2	9	[嘉義, 台灣, 中區, 嘉義, 三到, 知更鳥巢餅茶, 嘉義, 便硯, 湯熱開]
2	10	[嘉義, 嘉義, 中正]
2	13	[小英, 杯盤, 葡萄柚]
2	14	[柳橙綠茶, 柳橙綠, 柳橙綠, 柳橙綠]
2	16	[蘇打茶, 太硬]

### ● 轉為 DTM

轉為DTM (5) 參數有做更動，建議重新執行

參數設定

Input - 68

任務結果

保留詞彙

以換行符號區隔，e.g.  
國立中山大學  
西子灣  
壽山...

最多篩選詞彙數量

200

三

轉為DTM

(5)

參數有做更動，建議重新執行

參數設定


Input - 68

任務結果

統計資訊


202

字數



22025

文章數



任務結果

Show10entries

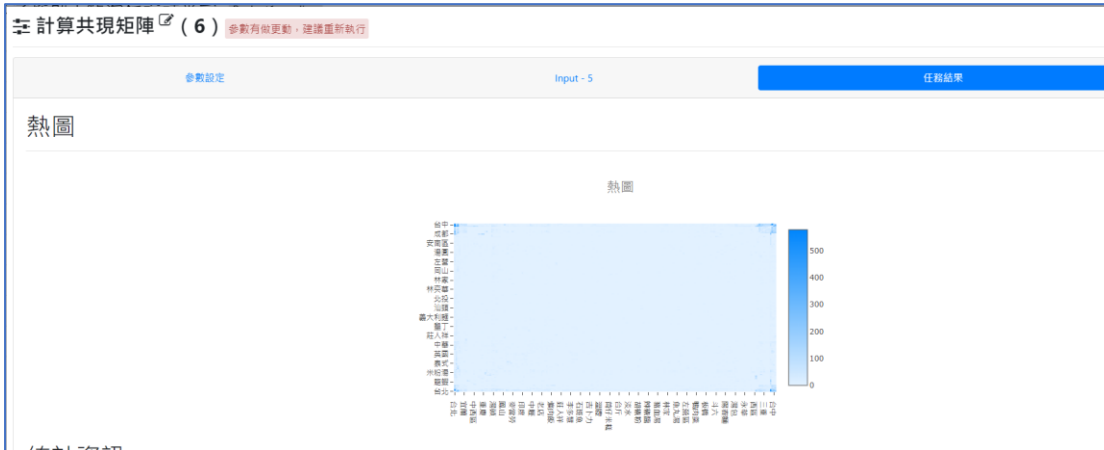
Search:

system_id	sentence_idx	三峽	三民	三民區	三重	上海	中和	中壢	中山	中山路	中正	中華	中西區	九份	亞洲	例確診	俄羅斯	信義區	內湖	刺皮辣椒	加拿大	北京	北投	北鎮	北門	南京	南區	南投	南橋	南港
1	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

### ● 計算共現矩陣

計算共現矩陣 (6) 參數有做更動，建議重新執行

參數設定	Input - 5	任務結果
儲存更改		



## ● 關聯式文字雲

關聯式文字雲 (shiny) (7) 參數有做更動，建議重新執行

參數設定	Input - 5	任務結果
------	-----------	------

分群數 \*

10

迭代次數(最少250次)

1000

聚合演算法 \*

single

文字雲顯示字數 \*

200

距離計算公式 \*

euclidean

儲存更改

## ● 單中心網路圖(Shiny)

單中心網路圖 (Shiny) (8) 參數有做更動，建議重新執行

參數設定	Input - 6	任務結果
------	-----------	------

選擇關鍵字典檔案

原始檔案元件 \*

DTM \*

共線/相關性矩陣 \*

儲存更改

## ● 字詞網路圖(Shiny)

字詞網路圖 (Shiny) (9) 參數有做更動，建議重新執行

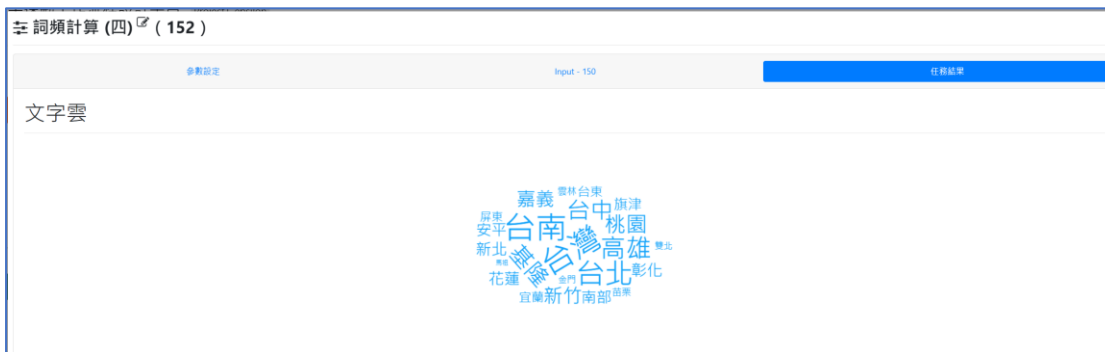
參數設定	Input - 6	任務結果
------	-----------	------

#### D. 視覺化的分析結果與解釋

## ■ 小吃詞頻文字雲



## ■ 地名詞頻文字雲

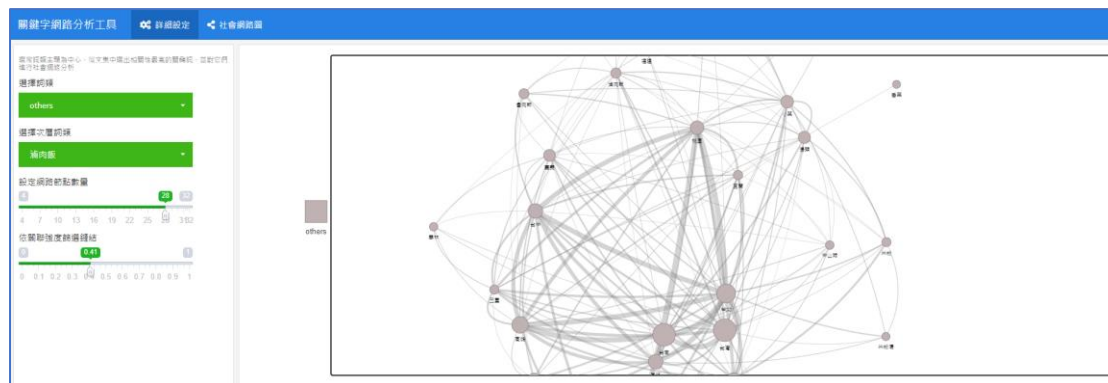


## ■ 關聯式文字雲

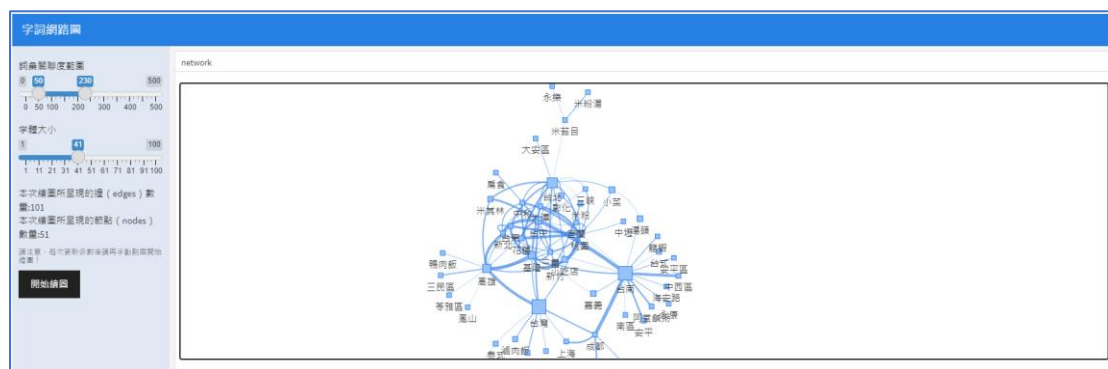


從圖中可明顯看出各地區的小吃有所不同，而台北、桃園、基隆、台中的小吃特性比較接近，而台南、新竹、高雄有獨立的群組關係，滷肉飯與魯肉飯與湯頭、小菜相關，可很明顯看出只要談到米粉都會和新竹掛勾，高雄與台南相似度較高。

## ■ 單中心網路圖

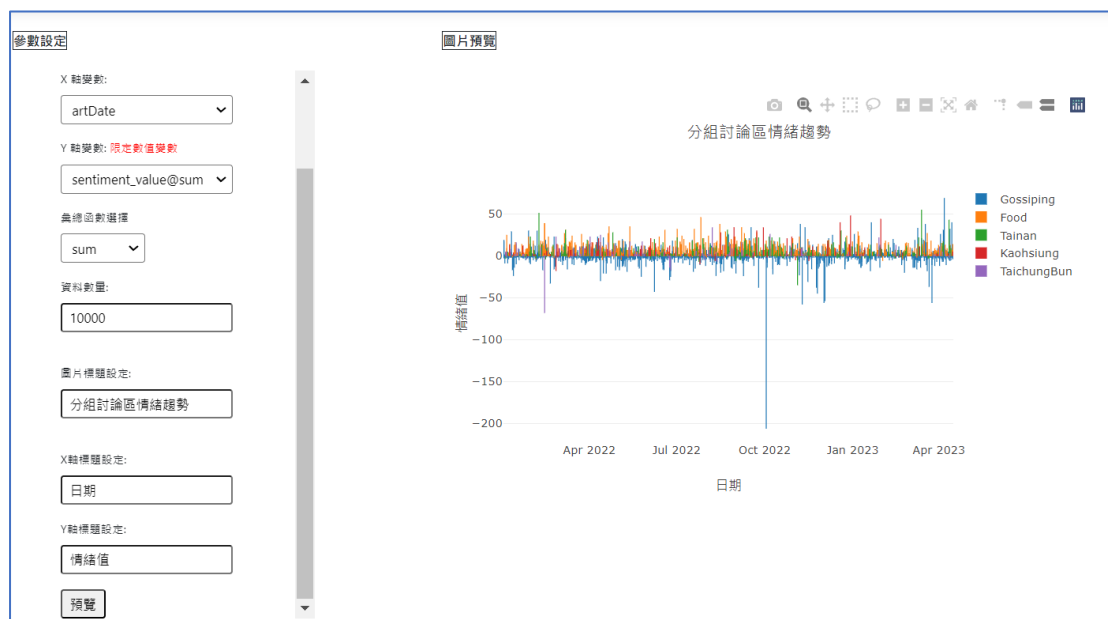


## ■ 字詞網路圖



從圖中可以明顯看出，各地區詳細的店名、小吃聚集之處

## ■ 分組討論區情緒趨勢圖



從圖表趨勢可以看到八卦版討論小吃，情緒大部分都是負面情緒

## E. 結論

根據本研究的分析結果，我們得出以下結論：

由關聯式文字雲圖可見，台灣各地區討論的小吃有所不同。台北、桃園、基隆、台中等地區的小吃特性較為相似，而台南、新竹、高雄則形成了獨立的群組關係。這說明了台灣各地區的小吃具有地域性特色，反映了當地的飲食文化差異。

從字詞網路圖中可以看出，各地區詳細的小吃著重點可能集中在特定討論店名、路名或區域，顯示出民眾對於特定地區小吃的關注和討論熱度。

根據 NTUSD 情緒分析結果，八卦版討論小吃時，情緒大部分呈現負面情緒。在正面情緒詞中，「在地」的詞頻排名第五，顯示在地化是民眾關心的重點之一。

綜合以上結果，本研究顯示台灣各地區的小吃具有顯著的地理分布和口味特點。在地化是民眾關注的重要因素，對於提升小吃的受歡迎程度具有重要意義。台灣各地區的小吃文化在社群媒體上廣泛傳播，形成了豐富多樣的美食交流氛圍。

未來研究可以從多方面探討台灣小吃產業的發展，進一步豐富和拓展本研究的成果。透過對消費者心理、行為動機、社群媒體平台、營銷策略等多個層面的研究，有助於揭示台灣美食產業的潛在機會和挑戰，為政府、業者和相關利益方提供寶貴的洞察和建議。跨地區和國際比較研究將有助於提升台灣小吃的國際地位和影響力，促進文化交流和產業合作。