



社群媒體分析

PROJECT 1

指導老師：黃三益教授

第17組

組長：N104020011陳冠宏

組員：N104020022凌瑩琪

N104020025陳姿樺

N104020028姜孟嫻

目錄

1. 動機與分析目的
2. 資料集描述
3. 資料的分析過程
4. 資料前處理
5. 視覺化的分析結果與解釋
6. 結論

封關前最後民調！高雄市長對決，5家民調結果只有這家不一樣

閻乙樺

+ 追蹤

2018-11-13 22:01 183679 人氣

簡

f 分享 1,349

LINE

f

...



民調今（13）日封關，各家民調紛紛公布最新選情調查結果。其中最受關注的高雄市長選情，國民黨高雄市長候選人韓國瑜網路聲量飆高，成為矚目焦點。

《蘋果日報》、《聯合報》、《中國時報》、《TVBS》與《三立》5家民調報告結果之中，韓國瑜大多領先對手民進黨高雄市長候選人陳其邁，其中《聯合報》結果更是大幅領先17%；但《三立》的調查結果當中，陳其邁則是小幅領先韓國瑜3個百分比。

動機與分析目的

陳其邁 (邁邁)

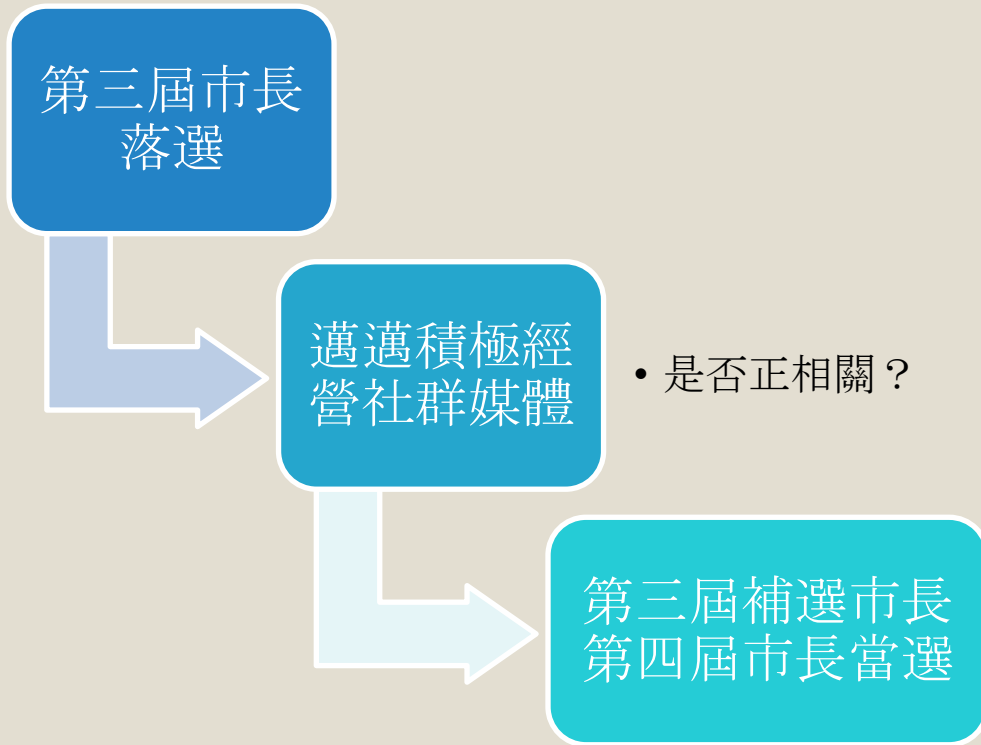
第三屆市長 落選

第三屆補選市長
當選

第四屆市長 當選

實際得票數相差
9%

動機與分析目的

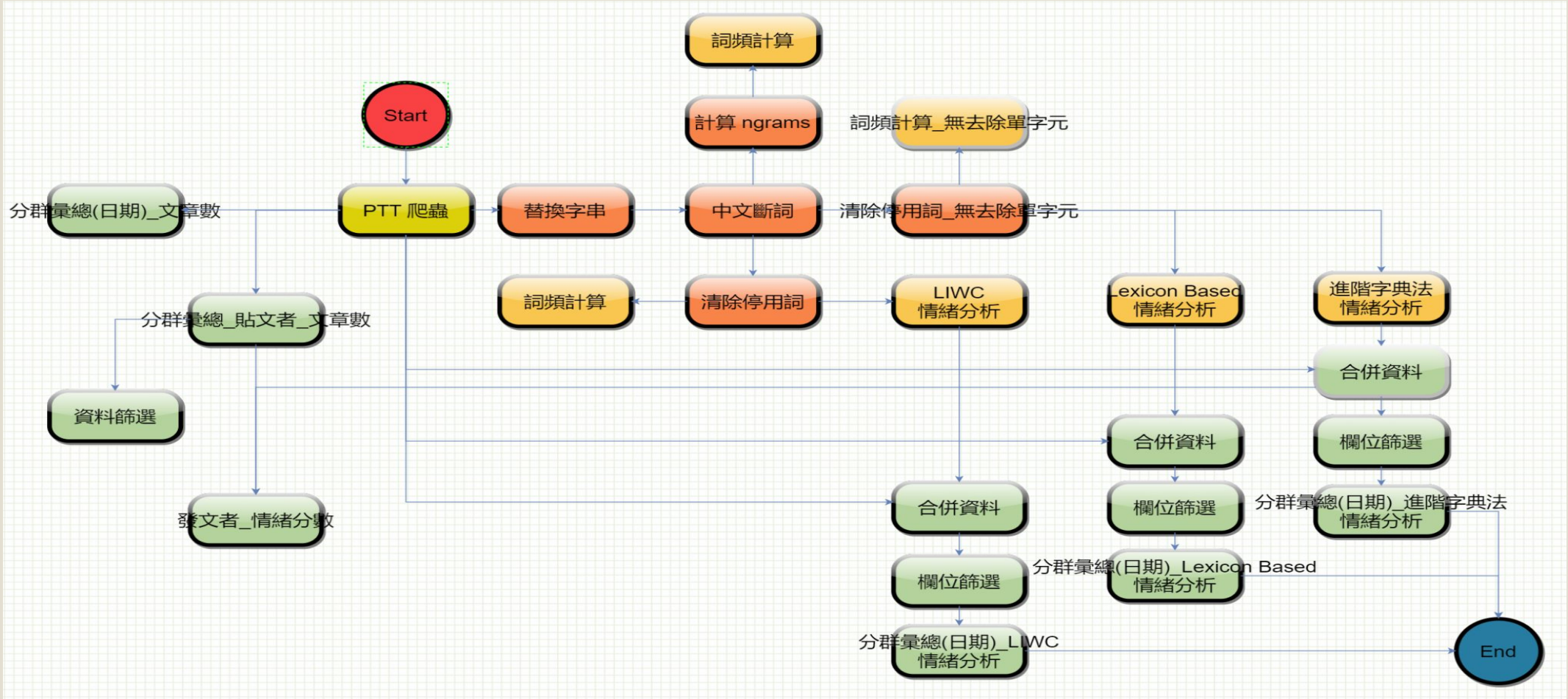


透過了解在PTT鄉民在第四屆選舉前對其的討論內容正負向情緒值，是否能藉此看出為後續當選跡象？
輔以新聞內容情緒分析進一步了解，新聞的內容是否會影響鄉民的評論走向。

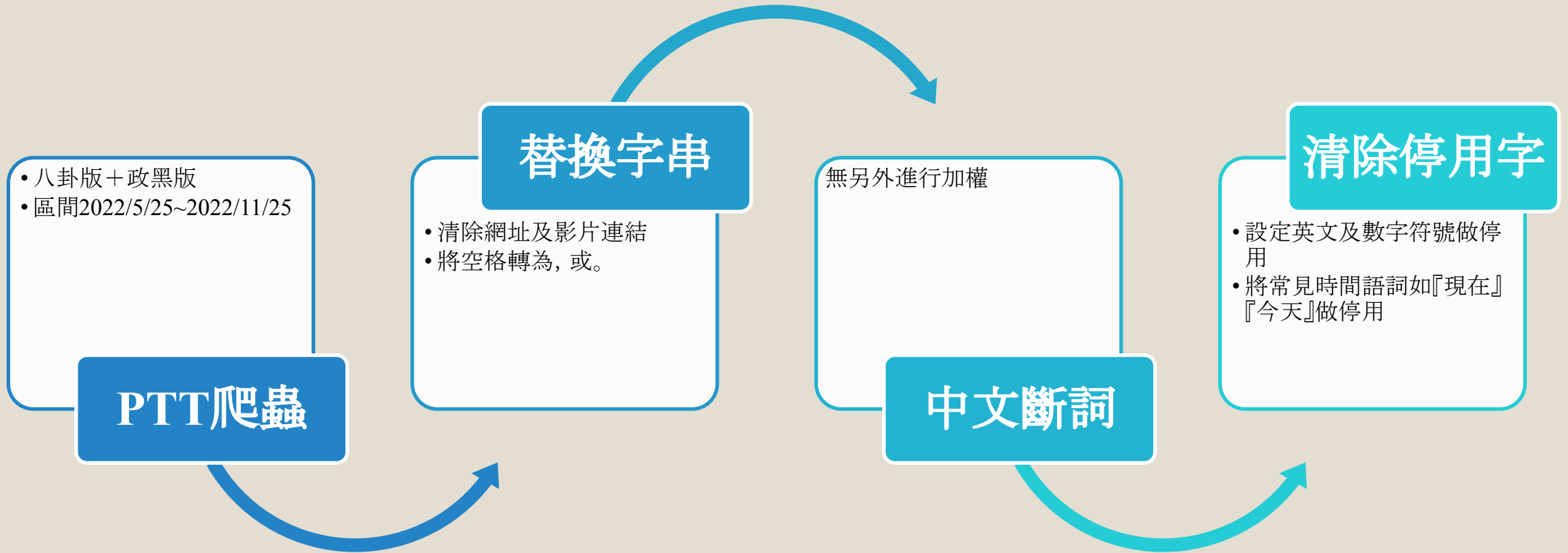
資料集描述

資料來源	PTT爬蟲(八卦+政黑)	聯合新聞(要聞+社會+地方)	東森新聞(全)
資料筆數	872	790	81
資料區間	2022/5/25~2022/11/25 (選前半年區段資料)		
關鍵字	陳其邁、邁邁		
排除關鍵字	顏寬恆、萊豬公投、顏清標、刺死、瑪莎拉蒂、盧秀燕、柯P、韓國瑜、韓總、陳時中、貴婦、塔羅占卜、羅智強、易經占卜、韓導、富邦人壽、內褲、基進黨、美腿甜心、許福明、論文、抽籤會、台北市長、中常會、台北市、新北市、彭文正、壞男人、馬英九、美豬公投、核四、林靜儀、屏東、轉錄網址		

資料的分析過程



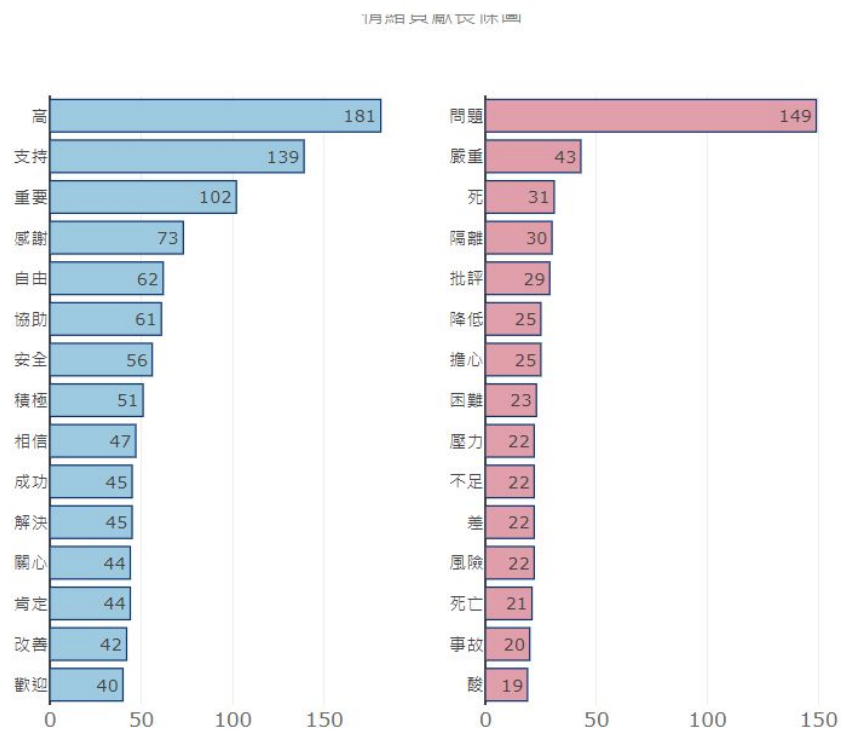
資料前處理



視覺化的分析結果與解釋

LIWC 情緒分析 (17)

參數設定



LIWC有加入單字詞，因此多了高、差及酸三個字於正向情緒分析中

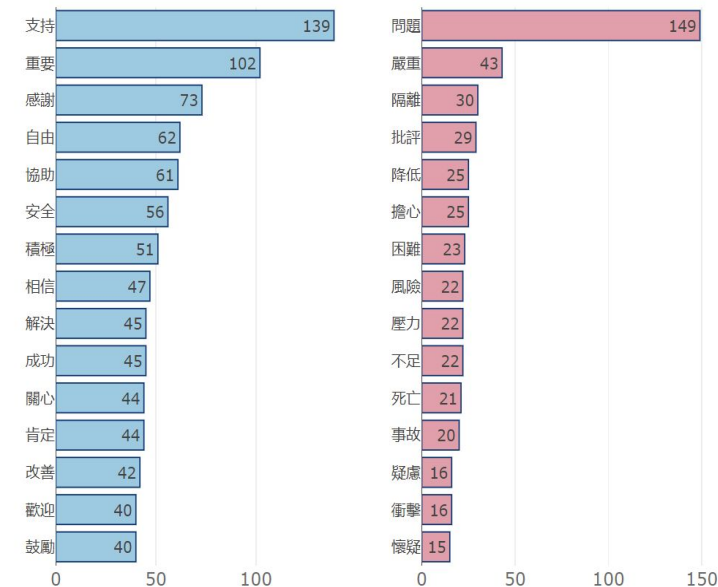
差為蠻確切的形容詞，屬於適合放入的詞，高及酸較突兀，我們重新將高雄、酸民兩個詞進行加權

LIWC 情緒分析 (17)

參數設定

情緒分佈

情緒貢獻長條圖



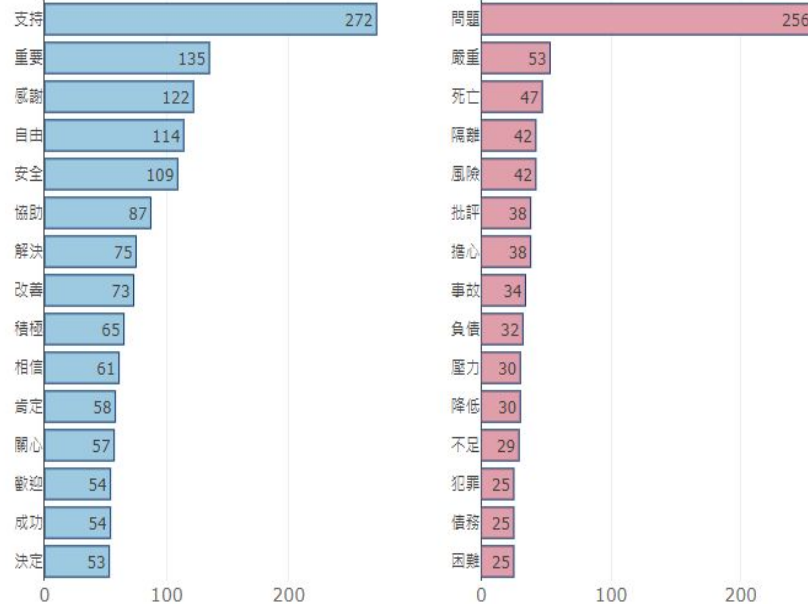
視覺化的分析結果與解釋

Lexicon Based 情緒分析 (22)

參數有做更動，建議重新執行

參數設定

情緒貢獻長條圖



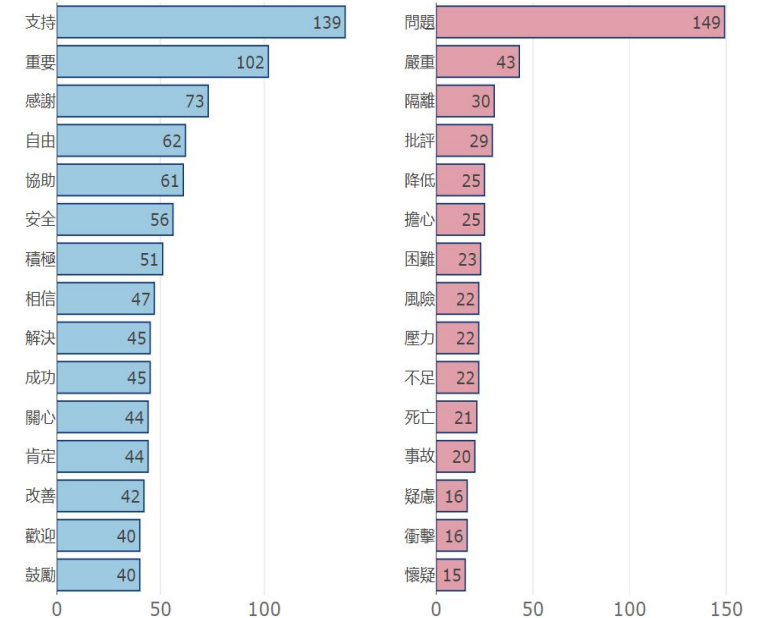
重新分權後針對
情緒分析用
LIWC及Lexicon
Based來進行比
對，發現兩邊字
詞差異不大。

LIWC 情緒分析 (17)

參數設定

情緒分佈

情緒貢獻長條圖

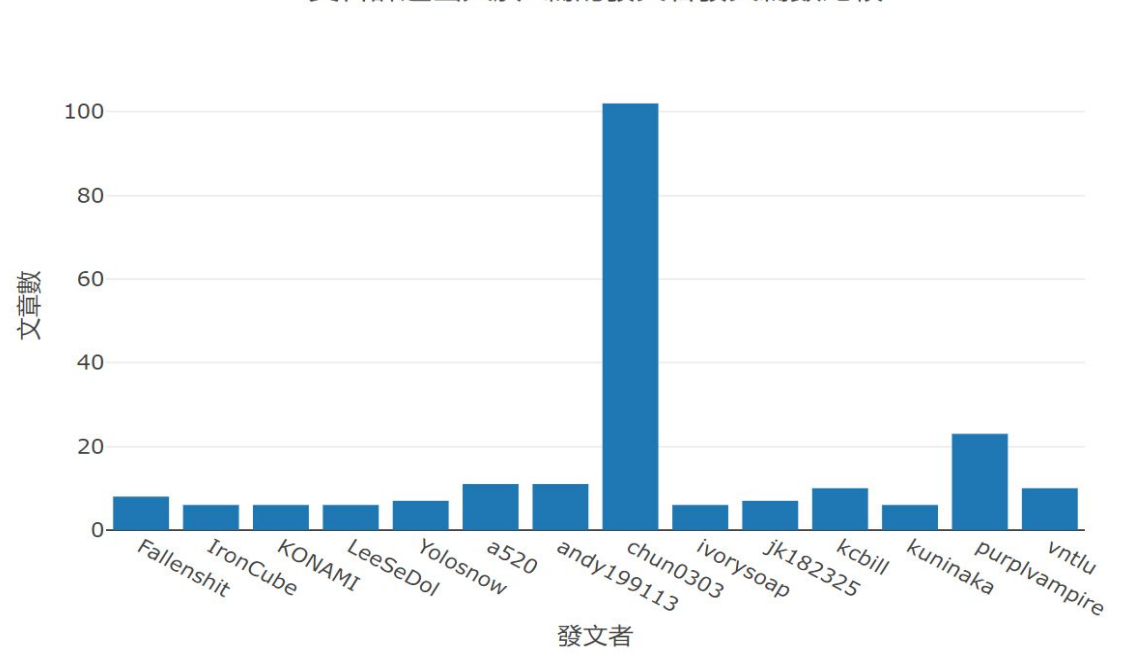


視覺化的分析結果與解釋

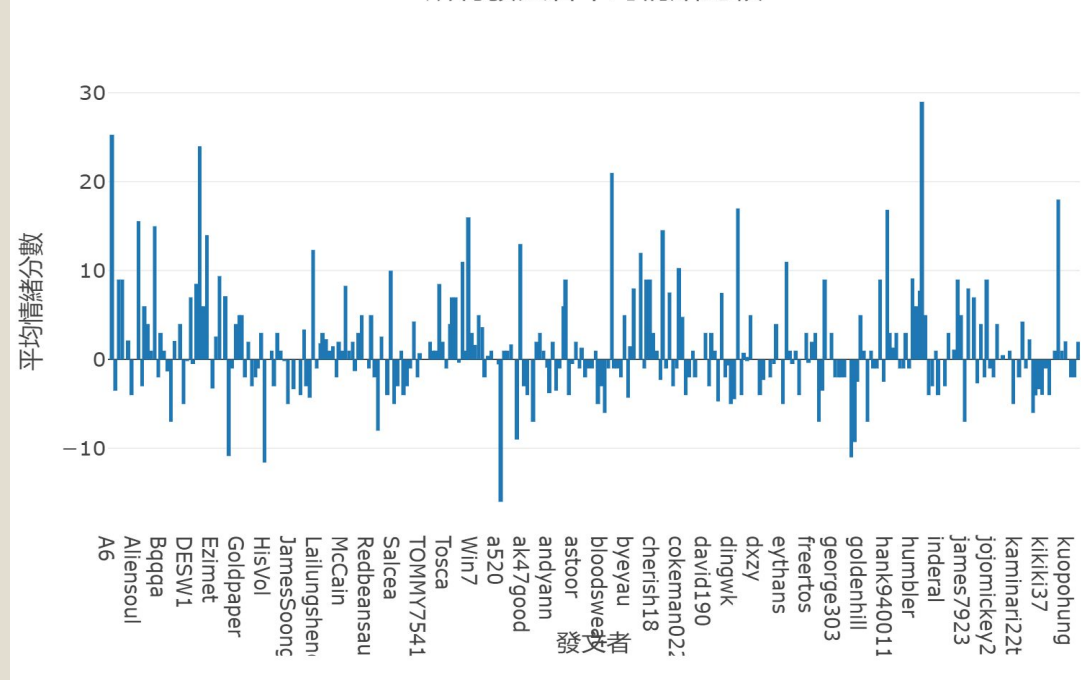
如下圖所示，我們發現有一位發文表現非常突出的發文者-chun0303，看了實際內容推測可能是新聞台的人，因為發文的標題都有個[新聞]的字眼出現

從下圖中發現到所有發文者的平均情緒以正面情緒居多

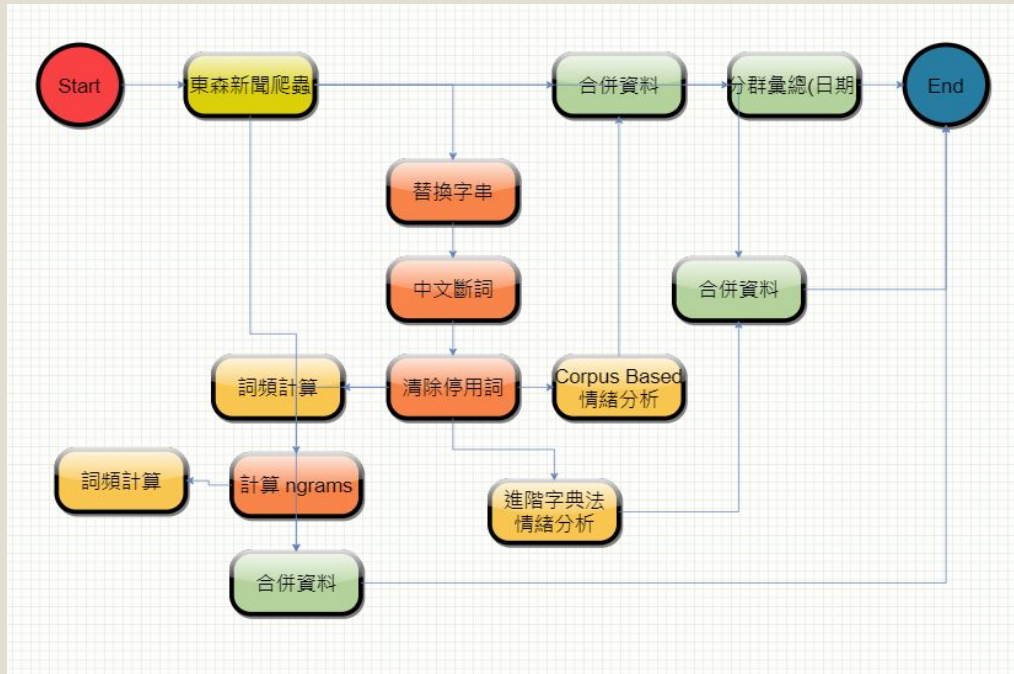
資料篩選出大於5篇的發文者發文篇數比較



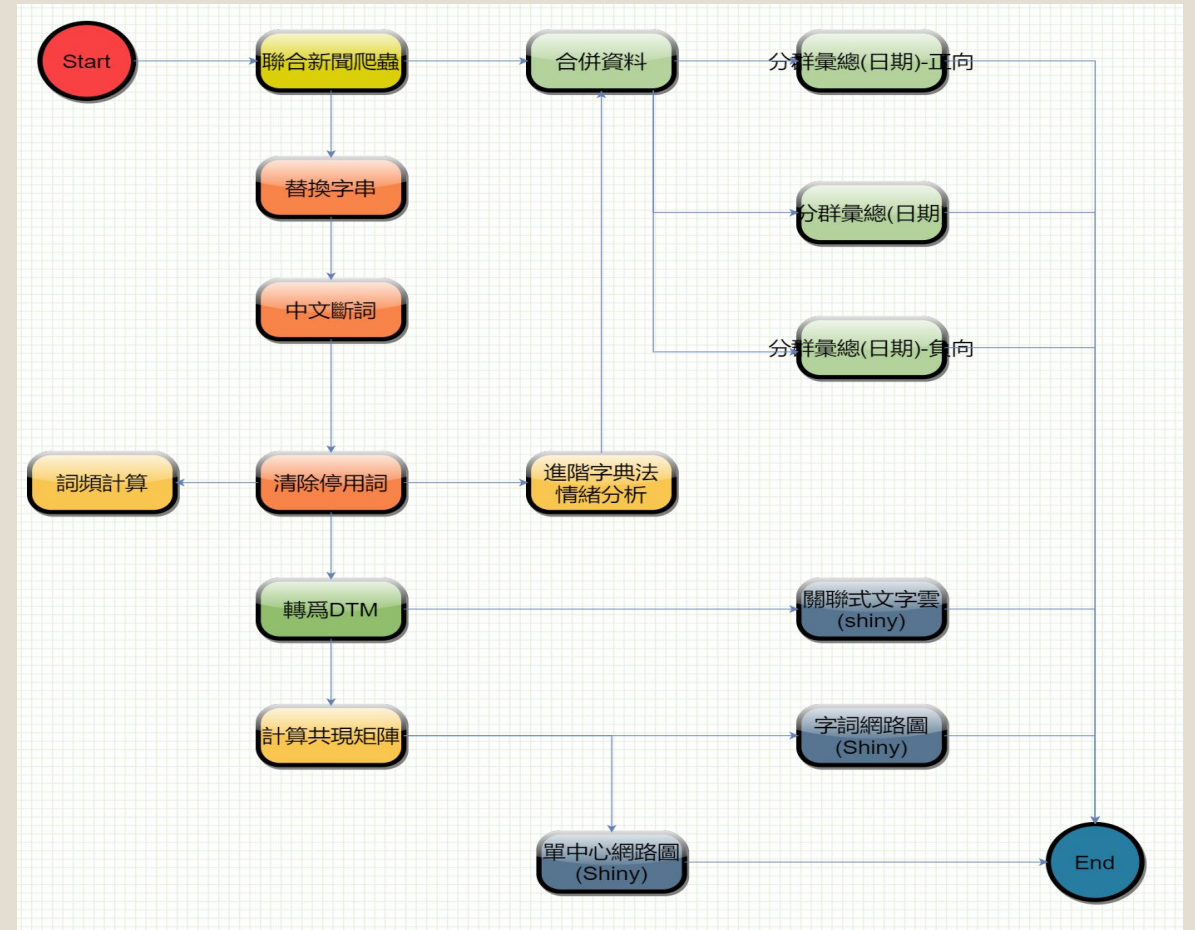
所有發文者平均情緒比較



資料分析過程__加入新聞

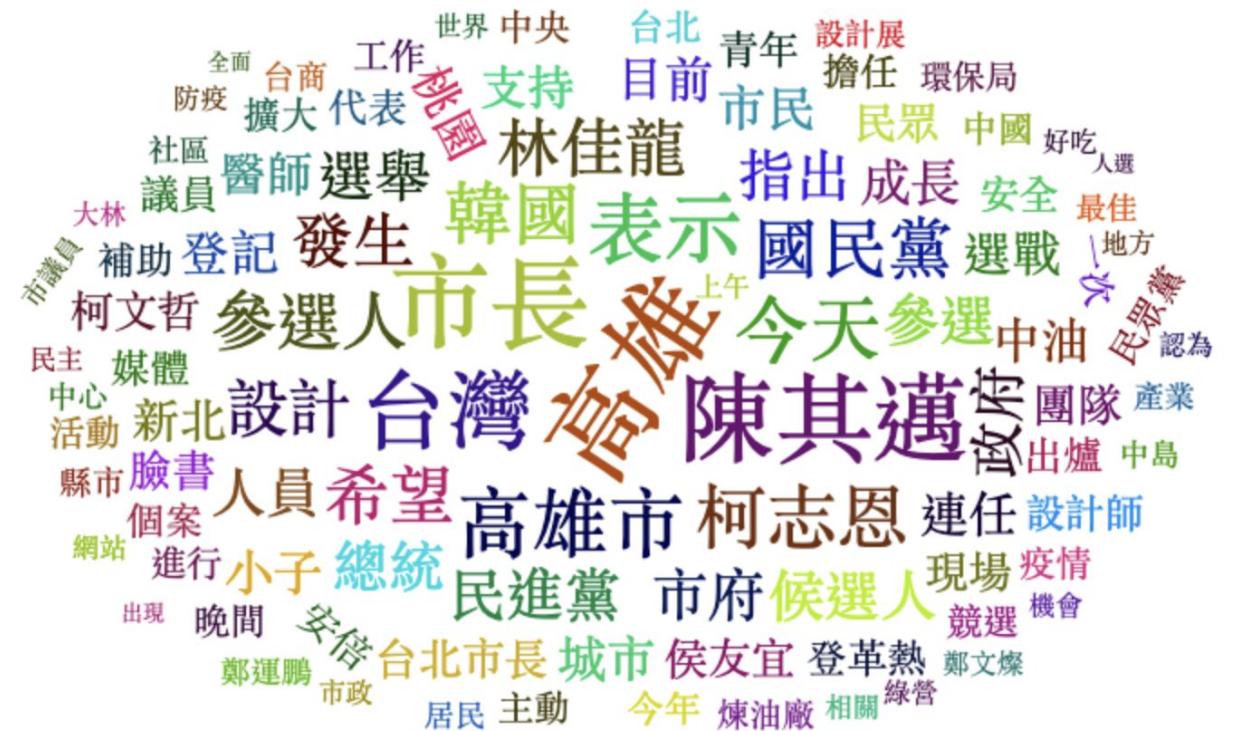
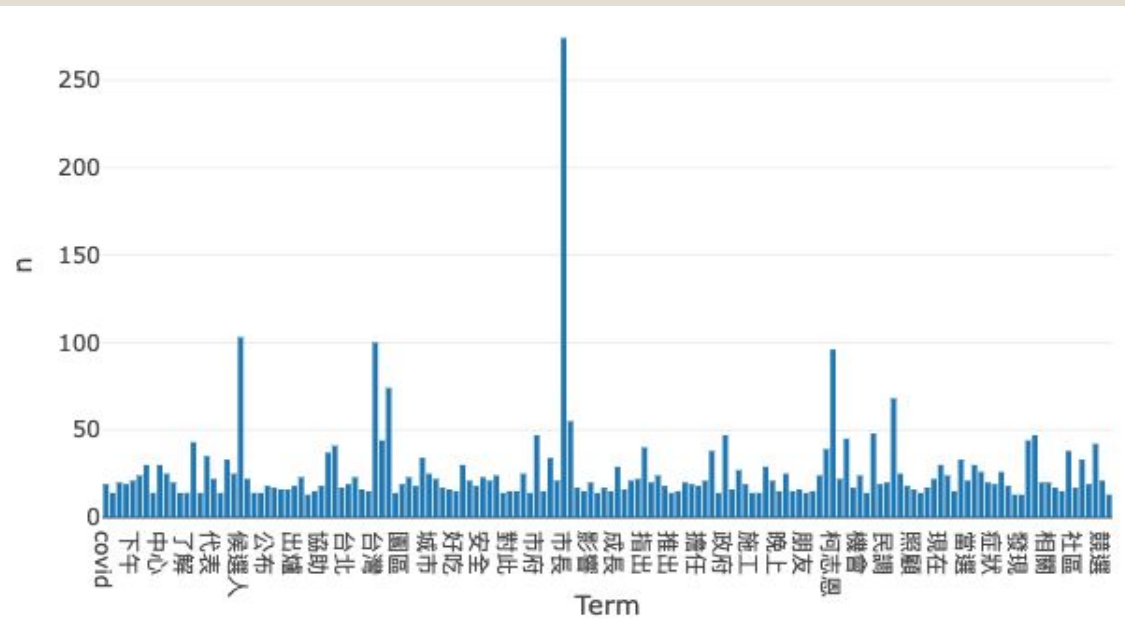


希望再進一步了解其與時事的相關性，
加入『東森新聞』、『聯合新聞』來進行交叉
比對



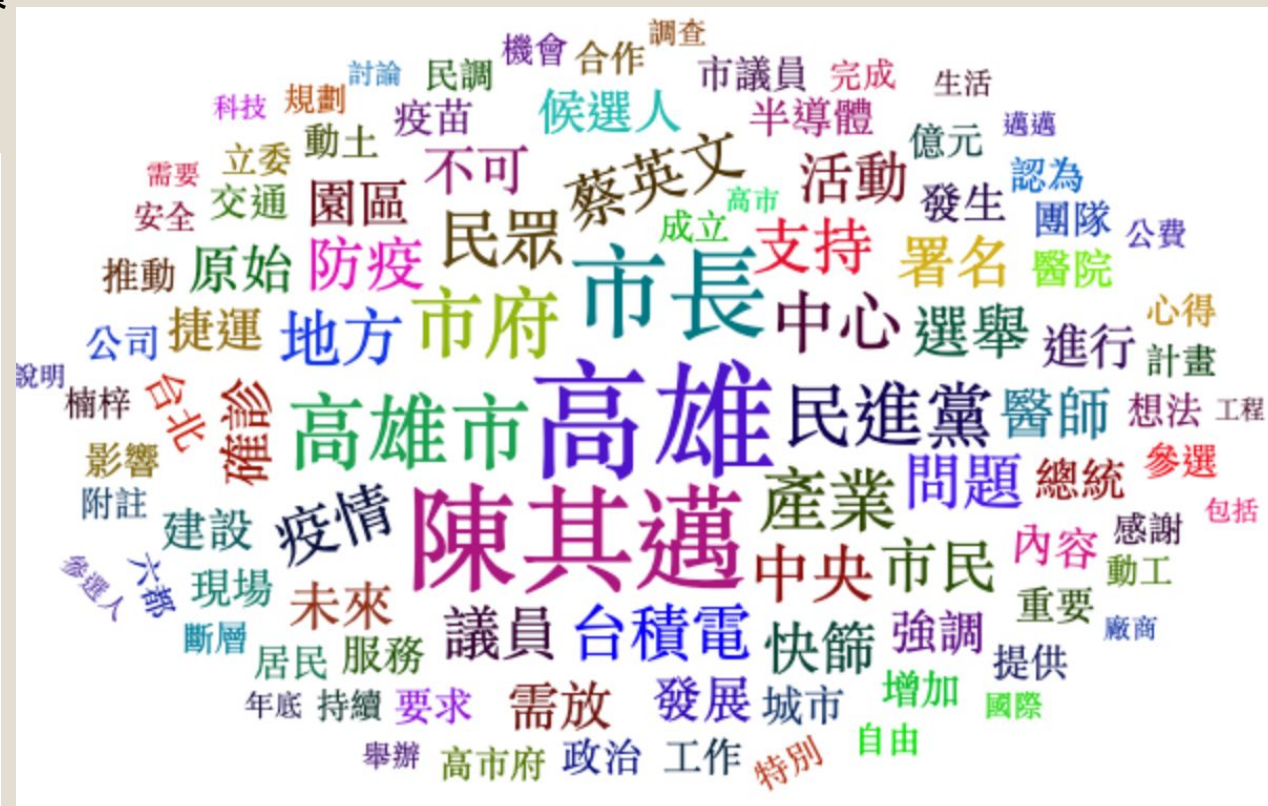
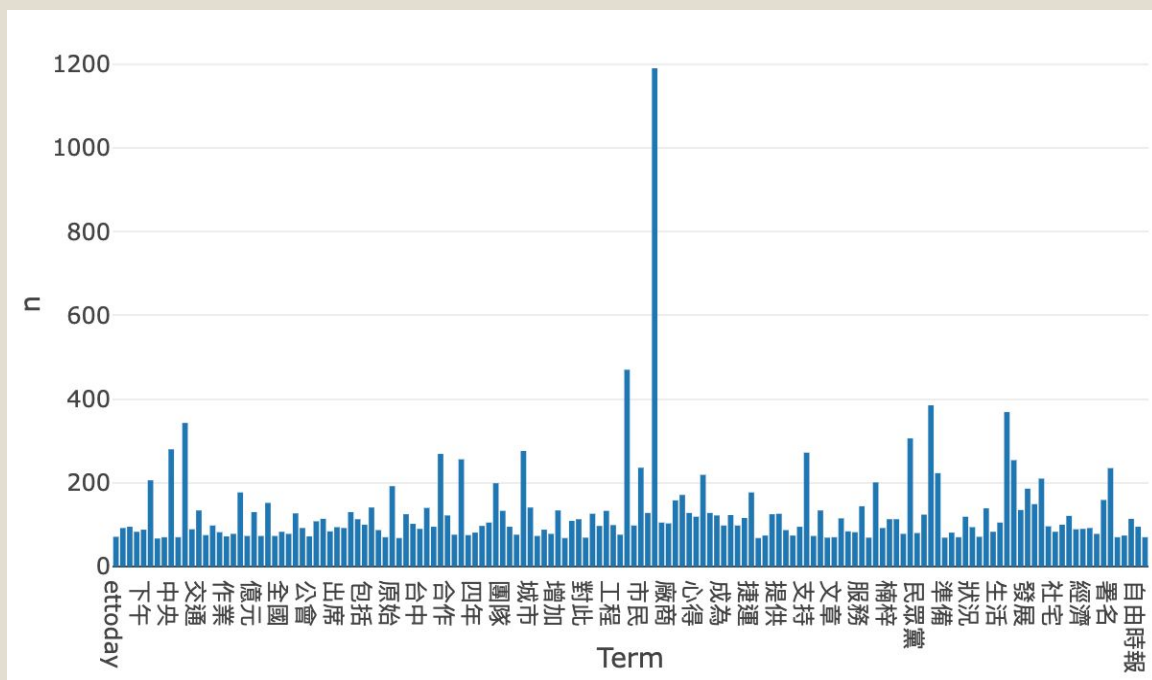
視覺化的分析結果與解釋

觀察5/25-11/25針對東森新聞分析詞頻計算結果，發現幾乎跟選舉有關，因此可以推測，PTT上的正向情緒，很有可能來自選舉時的支持者言論或是選舉相關新聞較多的關係。



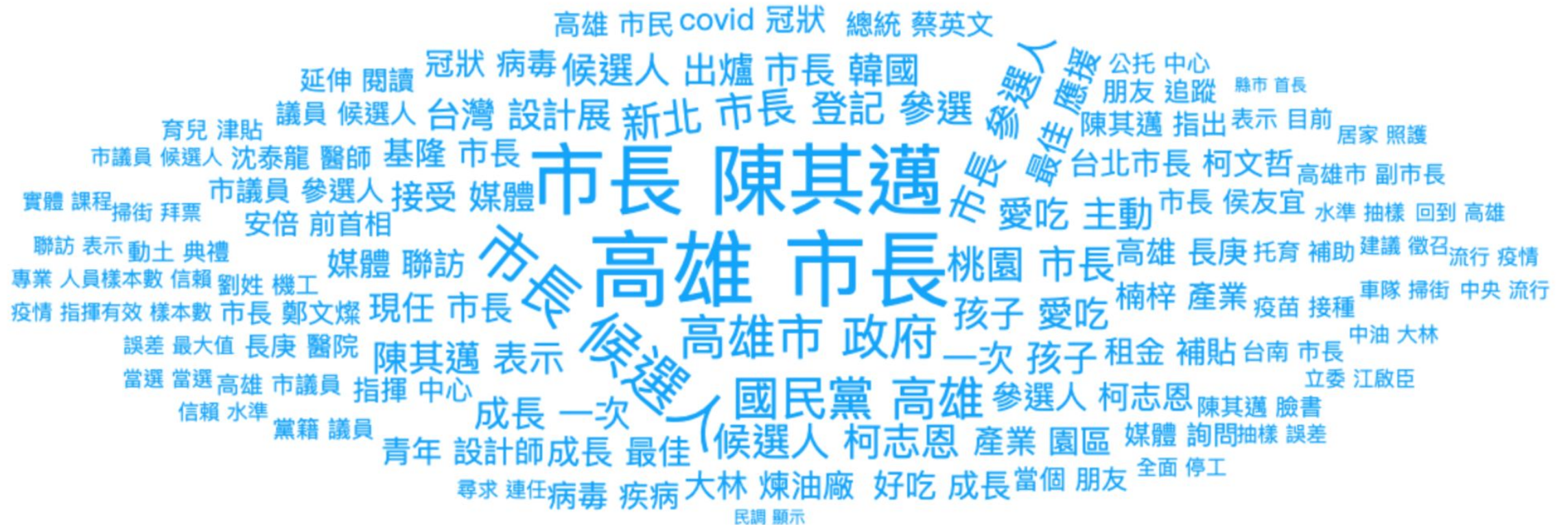
視覺化的分析結果與解釋

同期間PTT貼文分析詞頻計算結果，發現跟新聞的詞頻計算類似，跟選舉相關的字詞較多，另外還有一些跟最近熱門議題有關的字詞，下一步用ngram查看相關議題。

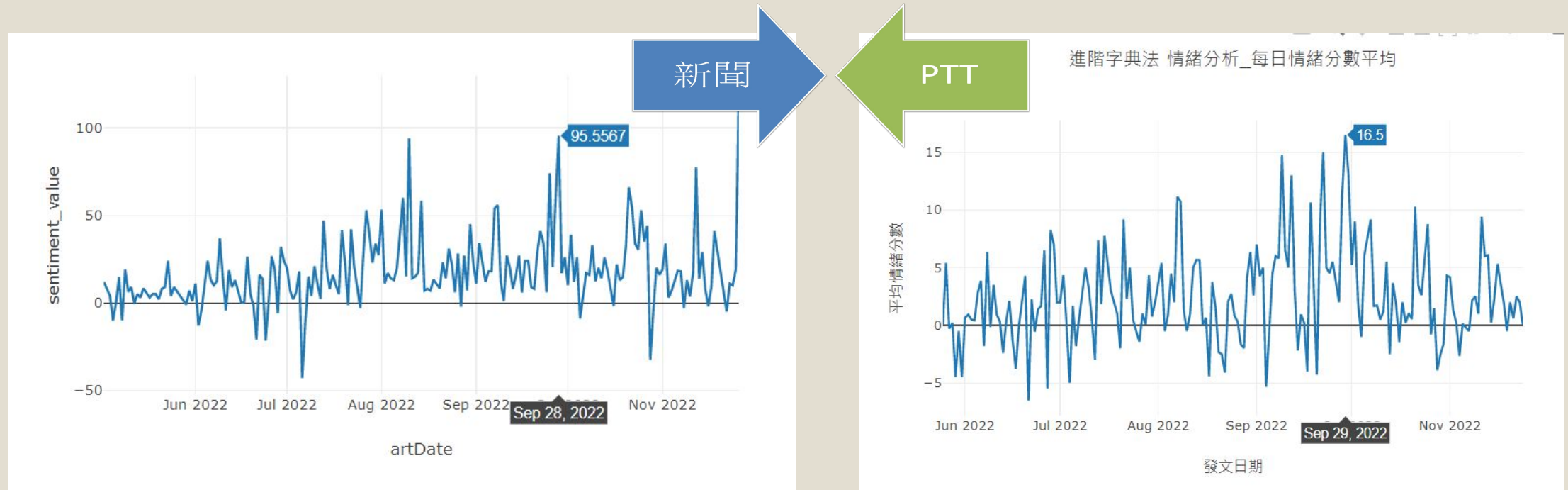


視覺化的分析結果與解釋

觀察ngram詞頻分析結果，可以看出一些地方上歷史熱門議題會在此時再度被提出檢視，例如疫苗接種、大林 煉油廠(曾發生大火)、楠梓 產業 園區(台積電建廠)等，由此也可看出市民在意的重點，候選人若有針對這些議題作為未來施政重點，也許可以爭取到選民的好感。



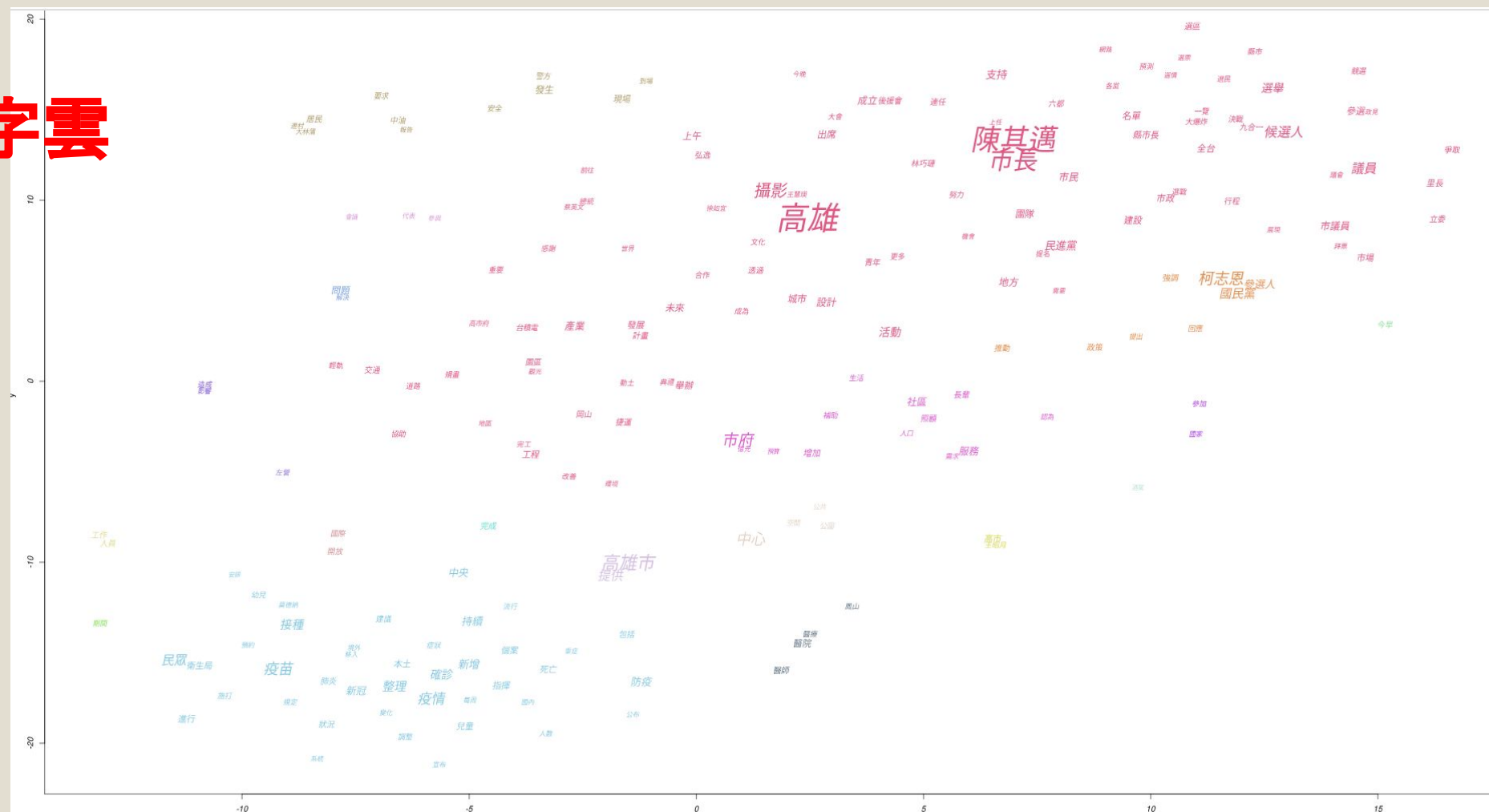
視覺化的分析結果與解釋



將聯合新聞網及PTT爬蟲情緒分群彙總繪圖比較，乍看之下沒有絕對的關聯，但是將PPT最高的日期為2022.9.29對照新聞情緒分析圖發現其於9.28也有一次高峰，推測新聞發布有影響鄉民討論度的可能性。

視覺化的分析結果與解釋

關聯式文字雲



視覺化的分析網

字詞共現圖

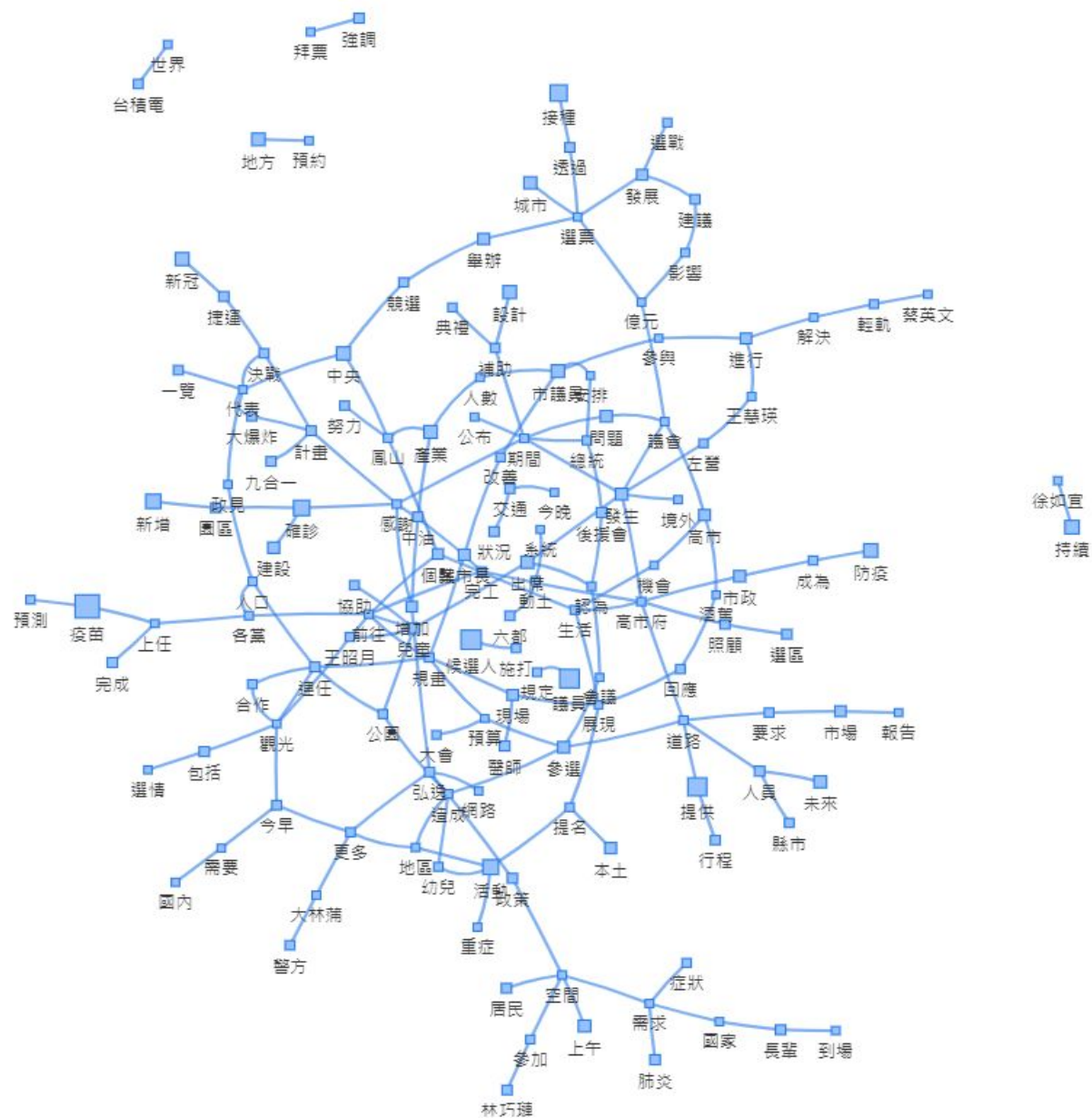
最常出現的幾個字

『疫苗』、『防疫』、『接種』、『新冠』、『確診』

『候選人』、『議員』、『活動』

『産業』

可以顯示出選民在這段區間最關心的議題就是選舉與防疫。



結論

擷取社群媒體選前半年的資料分析後發現社群媒體的輿論走向的確與選情有相關性，若輿論皆為正向則較傾向會當選。同時新聞的發布亦會影響輿論的熱度。

目前資料爬蟲來源中，本文幾乎都是以新聞為主，推測為本次分析結果正向詞彙偏多的主因。

本次報告時有嘗試以留言萃取功能但是因為資料筆數限制，無法取得分析結果，如果能分析留言，或許會有不同的結果。

針對PTT網軍的疑義，運用發文者的分析得出，的確是有發文數很高的人，如果就此下結論那的確是有水軍的，但問題是針對發文內容來看，其發表的內容主要是新聞，因此感覺只是轉貼相關文章的新聞工作者或者政治狂熱分享份子。



THANKS