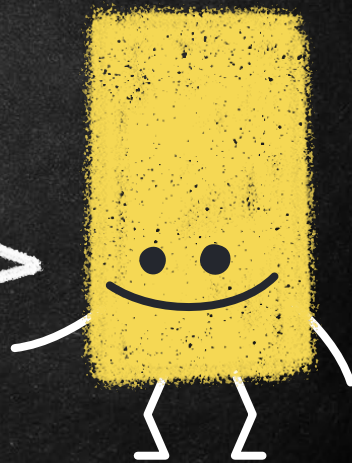
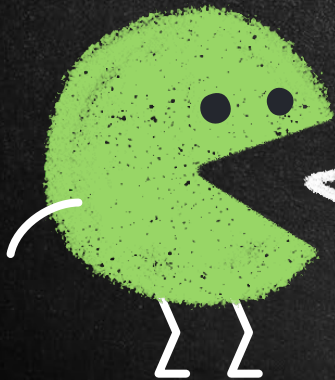


# 社群媒體分析 期中專案報告

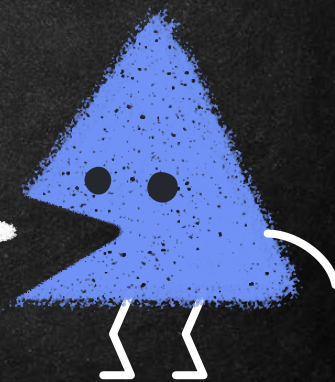
【ptt八卦版：台積電高雄設廠】

林志勳 N094220014 洪嘉駿 N094220020  
李明章 N094220023 陳雅詩 N094220026



# 大綱

- 研究動機與分析目的
- 資料集基本介紹
- 分析資料
- 分析結果視覺化解釋

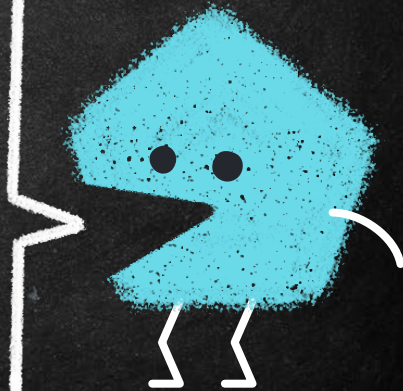
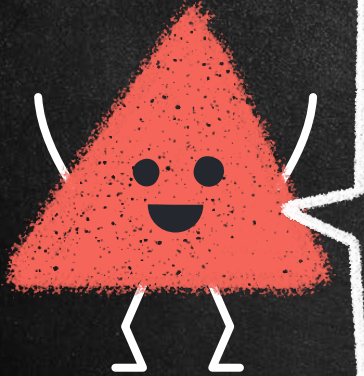




# 研究動機與分析目的

台積電預計在高雄設廠消息在今年9月初即在天下雜誌及ptt版網友間流傳，11月9日經台積電正式證實，因應市場需求，將在高雄楠梓的中油高雄煉油廠舊址設立7奈米、28奈米製程的晶圓廠，預計2022年動工、2024年投入生產。

在高雄設廠正式確定的新聞發布之後，引起許多在新聞媒體及社群的討論聲浪，因此選定PTT八卦版作為分析資料集的主要來源，並探討社群使用者對於該事件的態度及看法。



# 資料集基本介紹

資料來源：

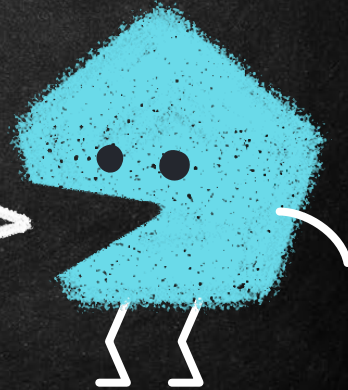
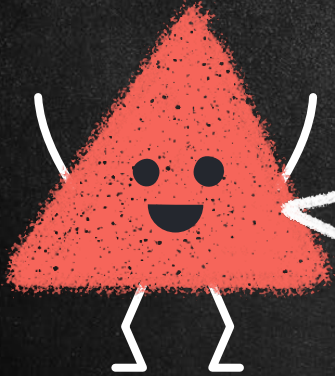
主要：ptt Gossip

其他：ptt的其他版如Kaohsiung、home\_sale、  
tech\_job、Stock

時間：2021/09/01 - 2021/11/20

關鍵字：台積電、高雄、設廠

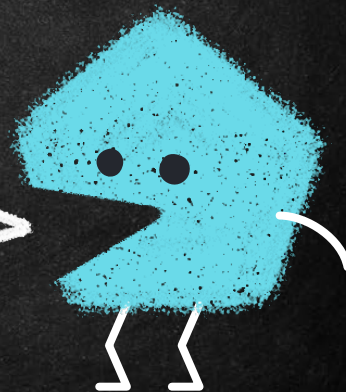
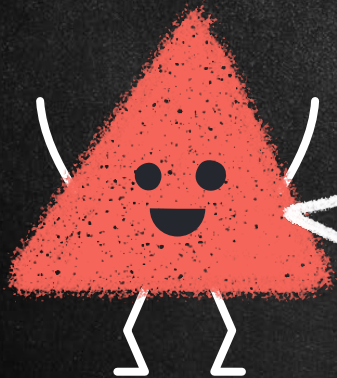
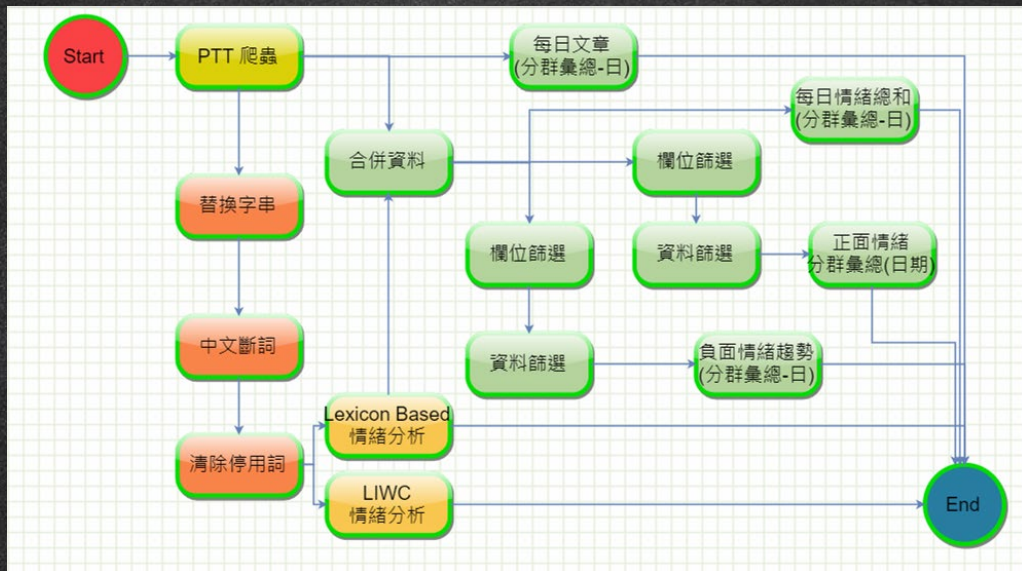
資料筆數：5786筆



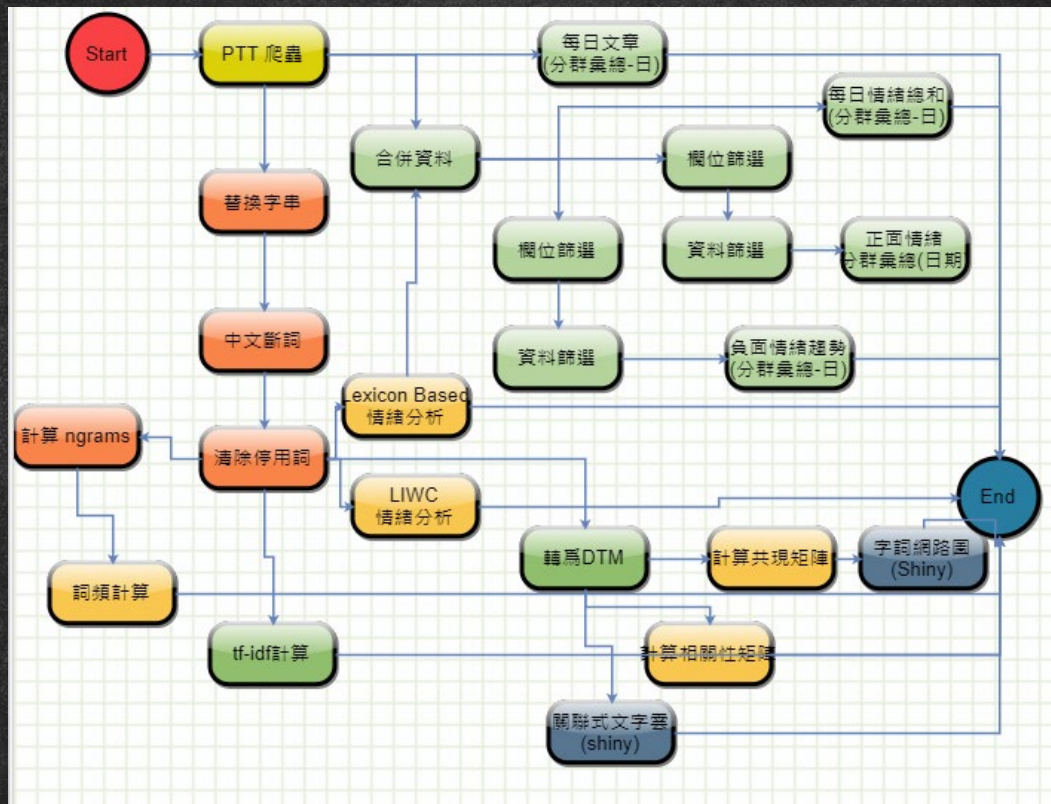


# 分析資料

## 使用工作流程平台 workflow



# 分析資料





# 資料前處理 替換字串

去除換行符號、網址、html標籤、底線、多餘的空格

替換字串 (7)

參數設定Input - 4任務結果

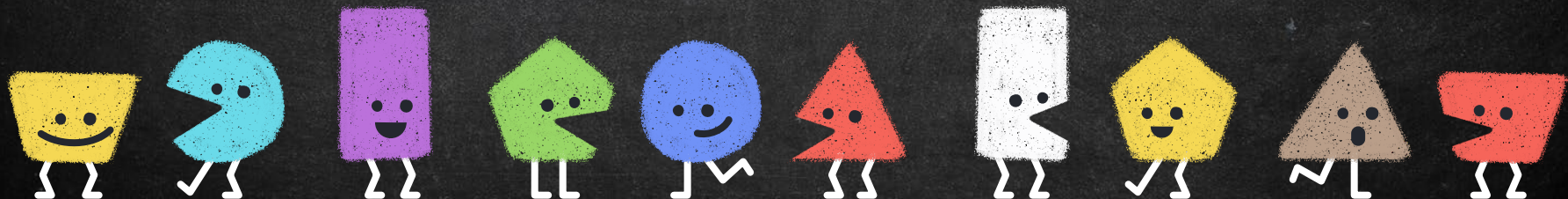
選擇處理欄位 \*  
artContent

替換字串設定 ⓘ  

```
\n>>  
https?:\\(www\\)?[a-zA-Z0-9@:~#%]{1,256}\\.[a-zA-Z0-9]{1,6}\\b([a-zA-Z0-9@:~#%&/=]*)>>  
\\W>>
```


選擇替換規則檔案 ⓘ  
dictionary\_example.csv

儲存更改

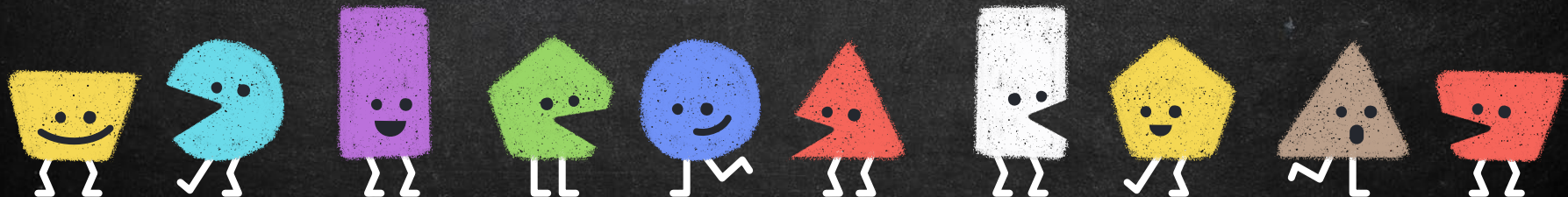


# 資料前處理 中文斷詞、清除停用字

進行中文斷詞、清除可能干擾分析結果的停用字

☰ 清除停用詞  ( 15 )

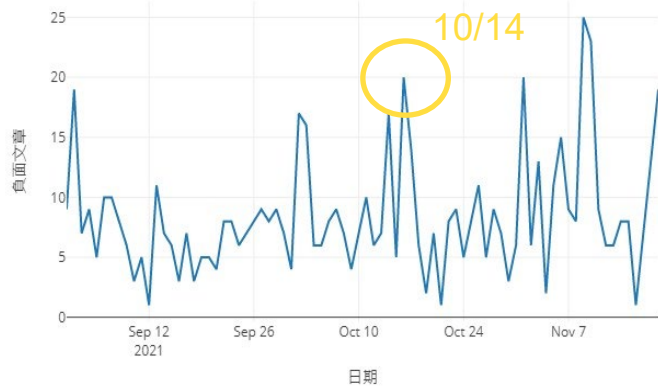
參數設定	Input - 28	任務結果
清除英文字母 *	清除數字 *	
<input type="text" value="是"/>	<input type="text" value="是"/>	
清除換行符號 *	清除特殊標點符號 *	
<input type="text" value="是"/>	<input type="text" value="是"/>	
清除html tag *	自定義停止詞	
<input type="text" value="是"/>	<div>你好 不要 問題 八卦 疫苗</div>	



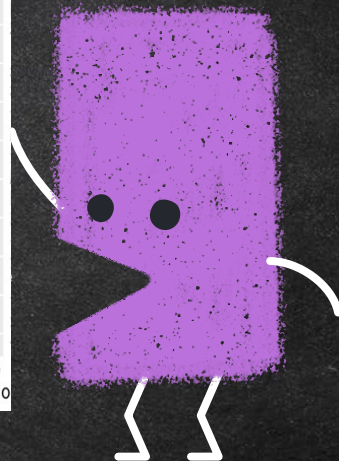
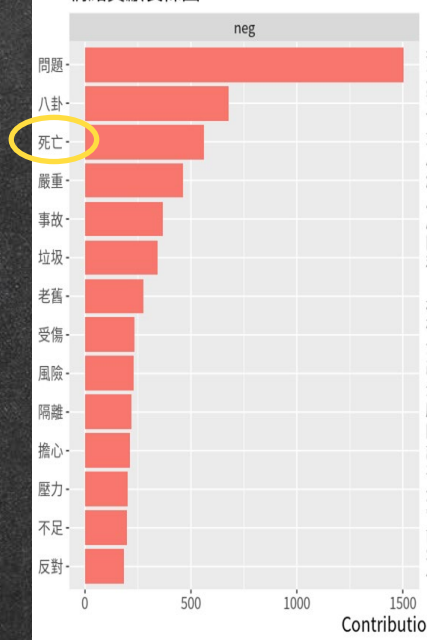


“

負面情緒趨勢(八卦版)



情緒貢獻長條圖







## 原始資料集

資料來源：PTT八卦版

時間：2021/09/01 - 2021/11/20

關鍵字：台積電、高雄、設廠

資料筆數：5786

## 去除關鍵字「高雄」 後之資料集

資料筆數：1218

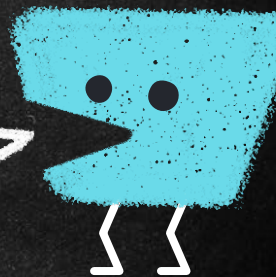
統計資訊

10

欄位數

1218

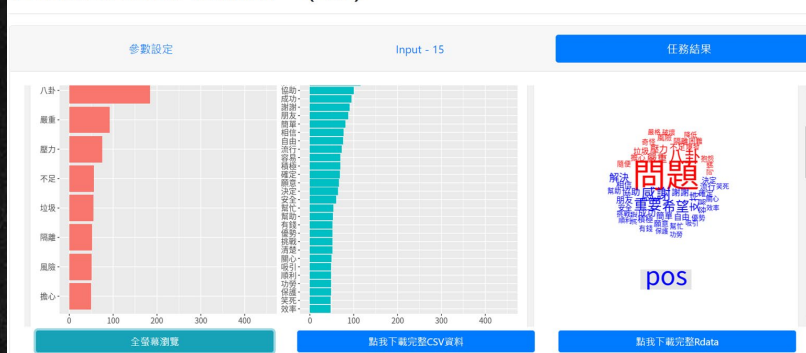
資料筆數



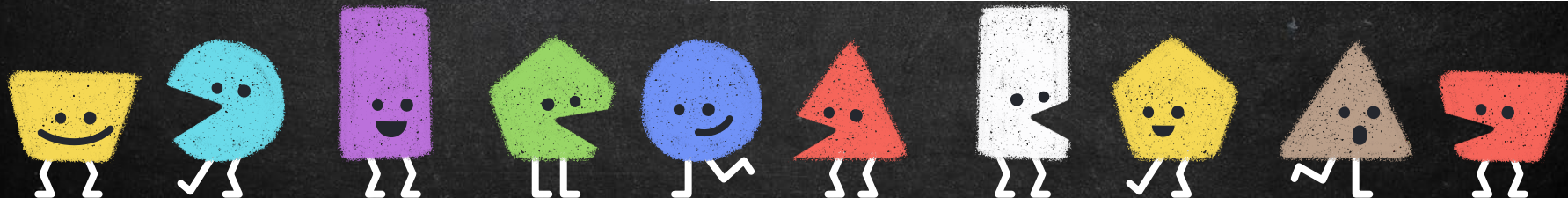
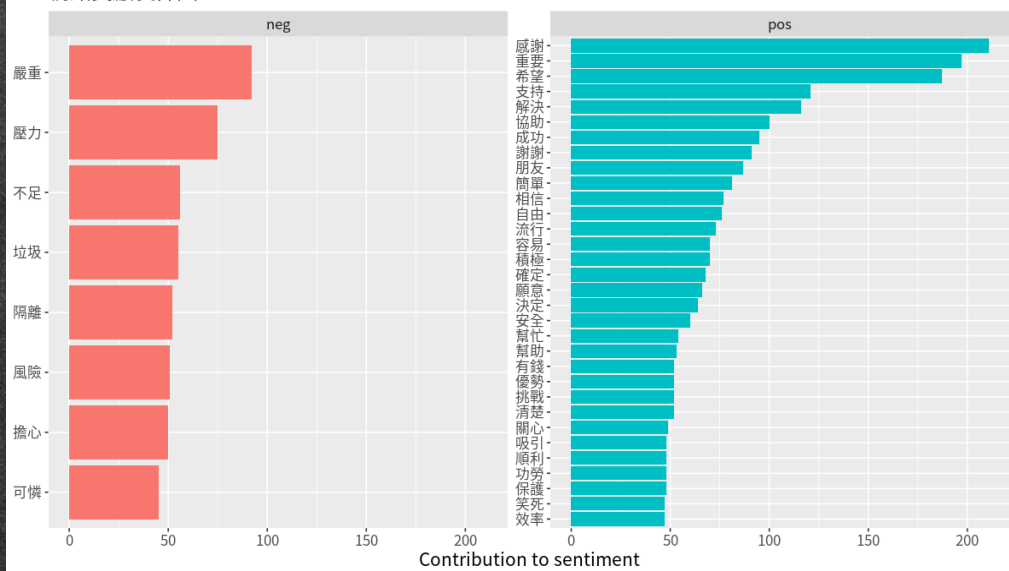
# Lexicon Based 情緒分析

將負面字詞中的“問題”、“八卦”等字做停用字去除後再做一次分析  
得到的負面字詞有嚴重、壓力、不足、垃圾，正面字詞有感謝、重要、希望、支持等

## Lexicon Based 情緒分析 (40)



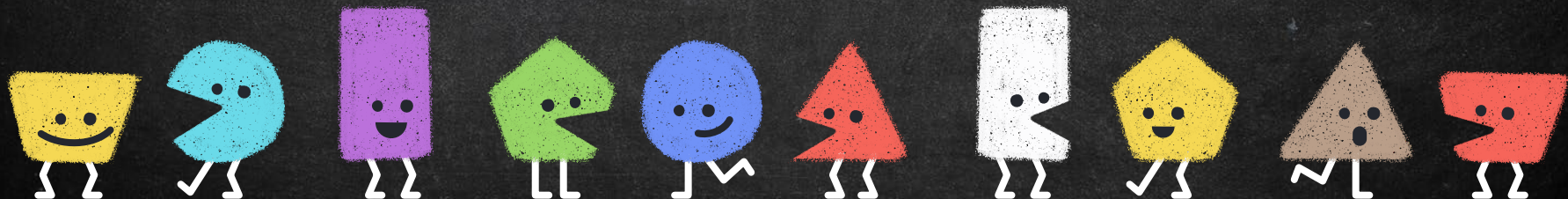
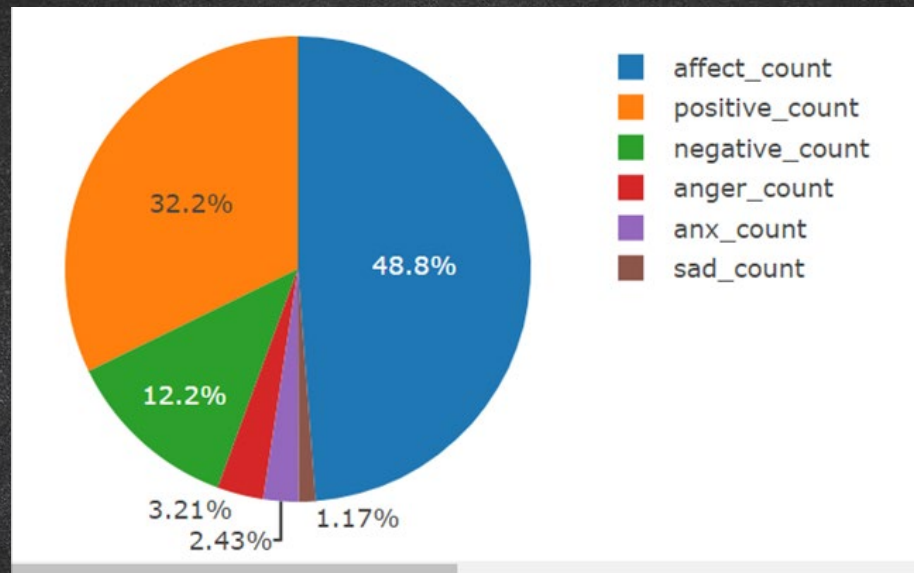
情緒貢獻長條圖





# LIWC 情緒分析

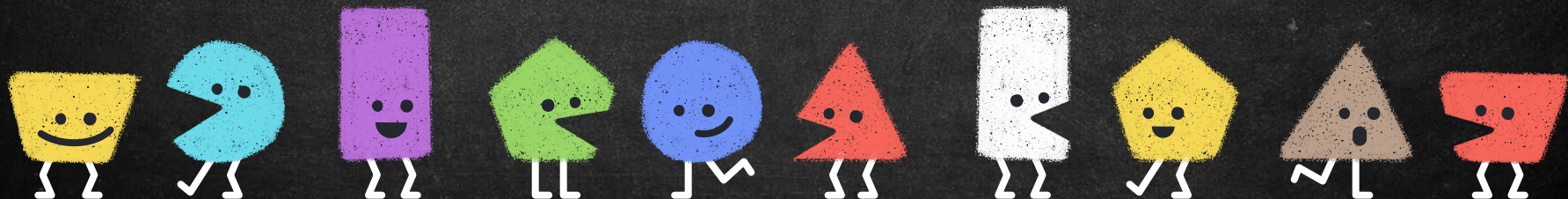
- 觀察實際的正負面文章數量，正面文章數量略多於負面文章數量
- 但是並沒有太過懸殊的明顯差距



# 將情緒分析與斷詞結果進行合併

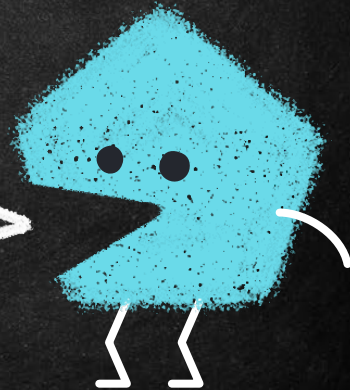
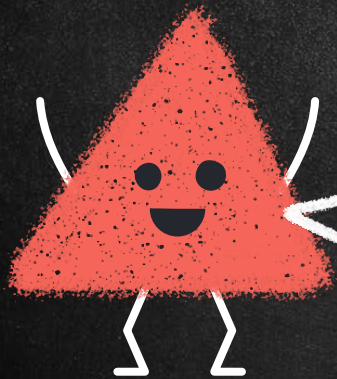
合併PTT八卦版和Lexicon Based情緒分析兩個任務結果的資料

進行欄位篩選: 參數設定: 選取artDate和 sentiment\_value , 任務結果列出分析資料期間  
 , 每日的sentiment\_value , 進行負面資料篩選(參數設定: \$sentiment\_value < 0)及正面  
 資料篩選(參數設定: \$sentiment\_value > 0)



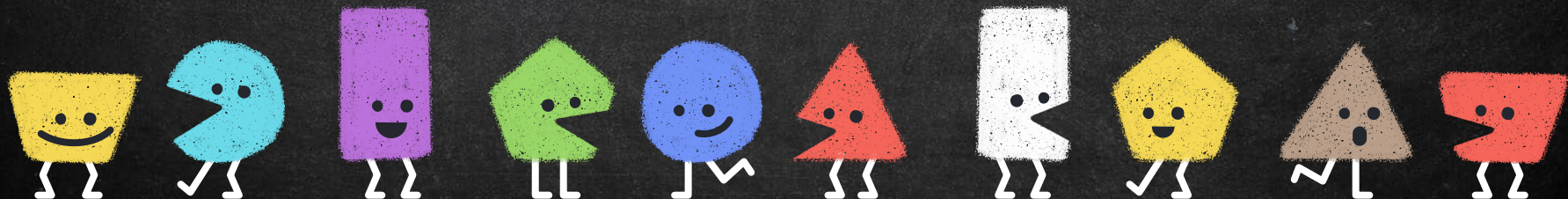


# 分析結果解釋



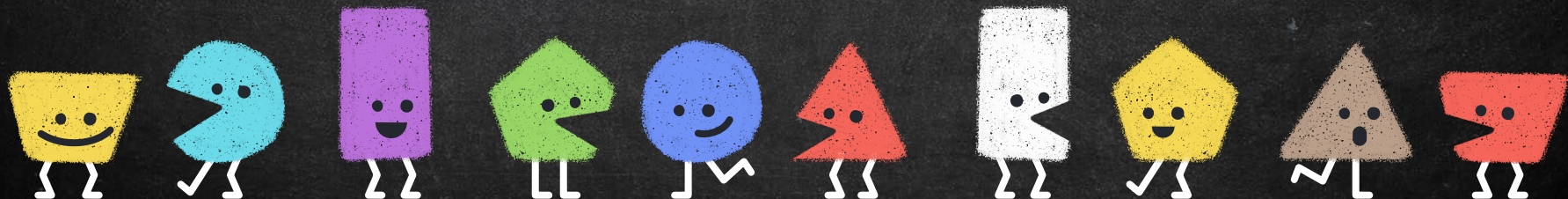
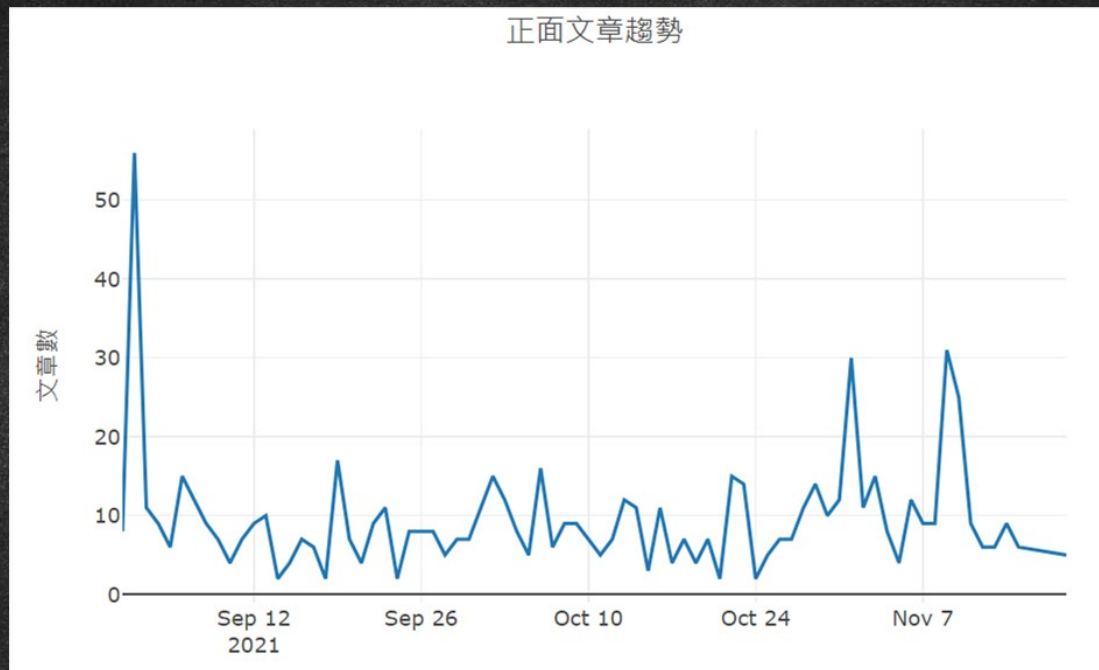
# 每日文章統計

八卦版有關台積電的討論文章數量最高的時間有兩個高峰，分別為**9月6日**(天下雜誌及網路新聞相關報導)及**11月9日**(台積電對外正式宣布證實高雄設廠訊息)

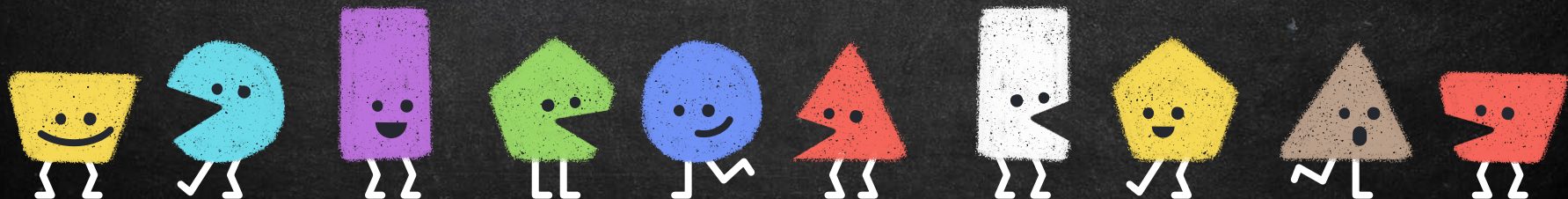
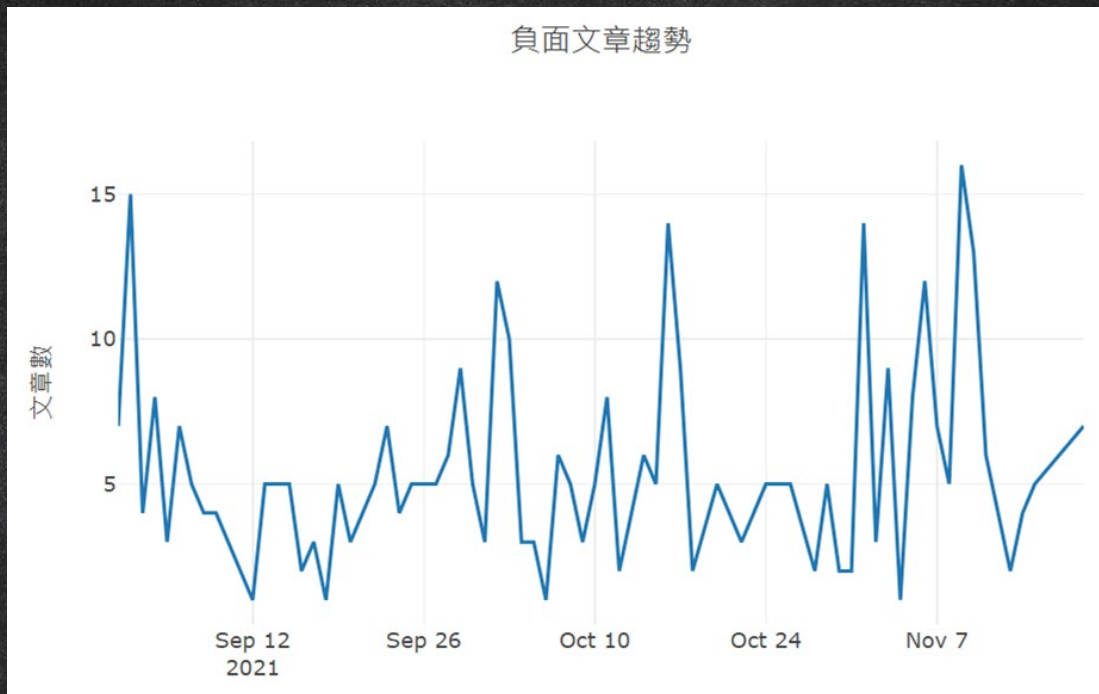




# 正面文章趨勢圖

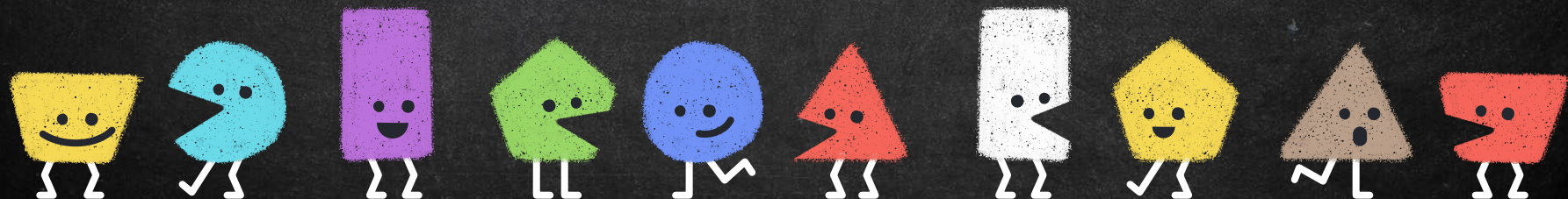


# 負面文章趨勢圖





# 詞頻計算 文字雲



## 使用其他與此主題相關的ptt版分析

# Kaohsiung 高雄版

分析主題為台積電高雄設廠，故推測在高雄版的討論度亦高

## home\_sale 房屋版

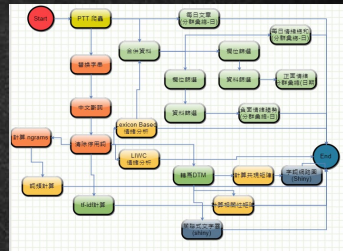
據社群及相關媒體報導，台積電高雄設廠，導致楠梓周邊地區房價攀升，故推測在房屋版亦應有相關討論

## Tech\_job 科技工作版

台積電設廠也將在高雄創造新的科技工作機會，故應也有相關討論

## Stock 股票版

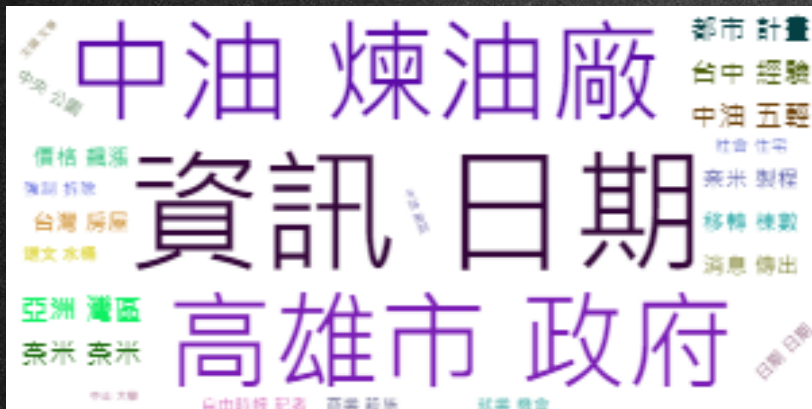
## 觀察台積電設廠消息在股票版是否有引起社群相關討論



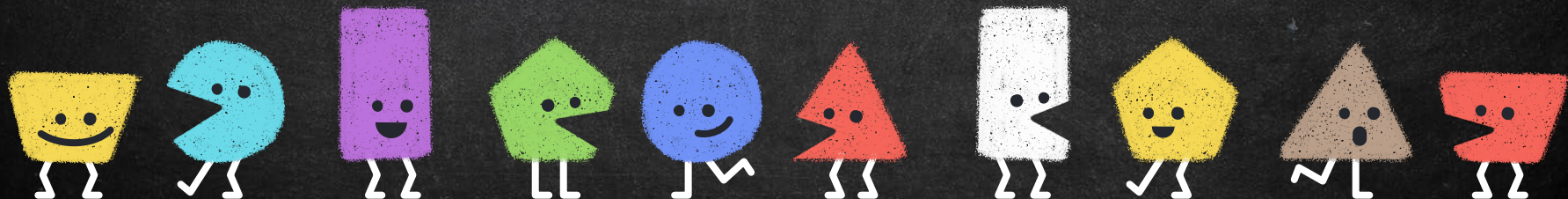
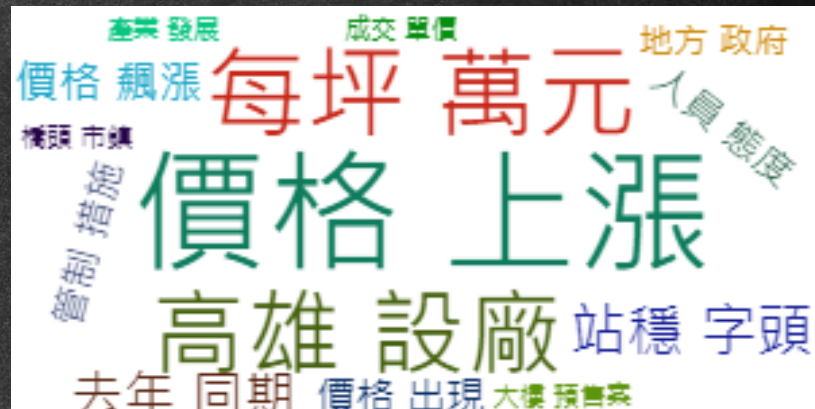


# 出現不同關鍵字的文章雲

Kaohsiung 高雄版

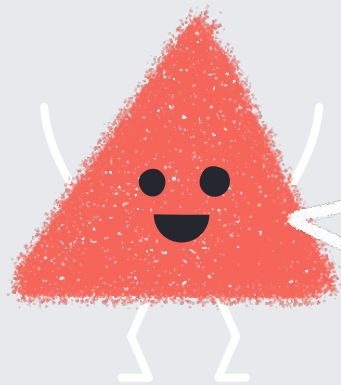


home\_sale 房屋版

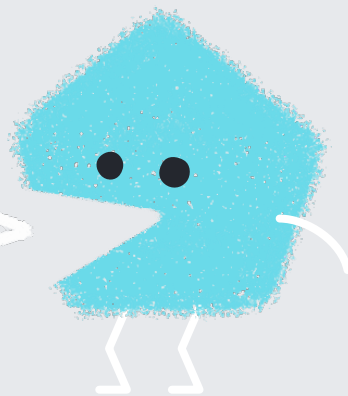


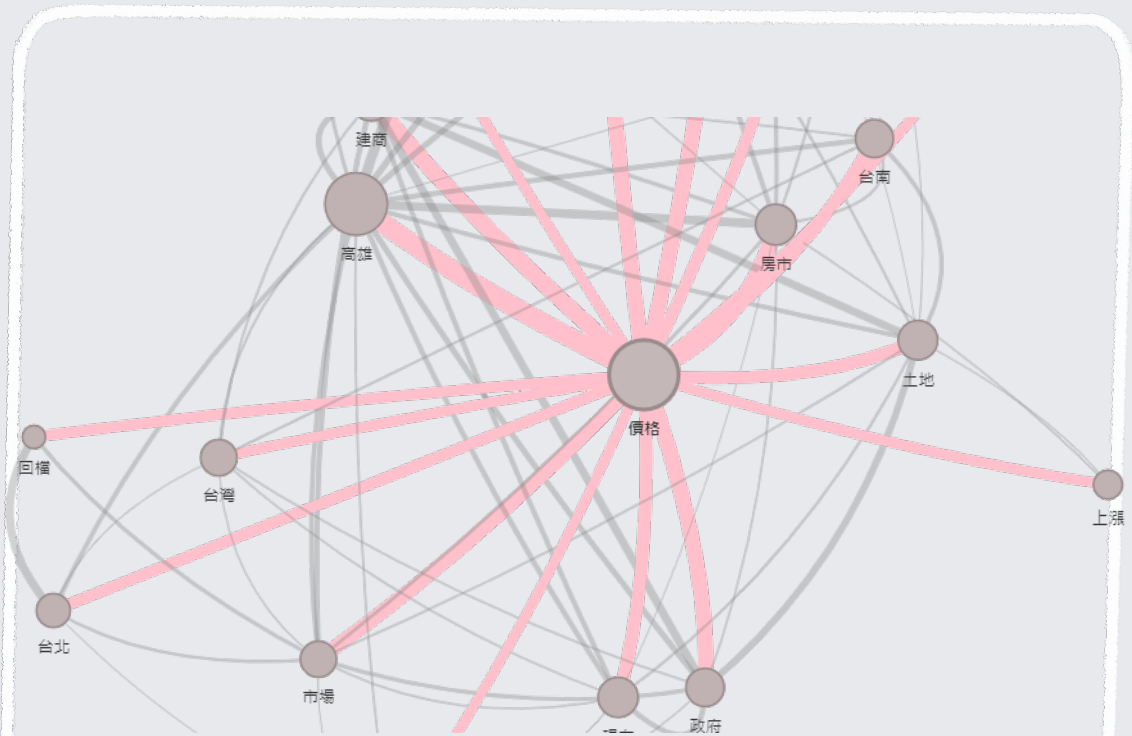






在home\_sale所產生的文字雲中，我們發現了出現「價格」、「飆漲」等關鍵字，透過TF-IDF對文本重點字詞進行加權之後，我們畫出了在home\_sale版裡針對本分析主題「台積電高雄設廠」的字詞網路圖及關連式文字雲，可以看出「房市」與「價格」兩個字詞的關聯程度很高，在本分析主題的資料集中出現次數也很高





## 字詞網路圖Shiny

Recent Co-Accurance (76): 房市

購買法拍屋容易出現屋況不佳 或得標後 需一筆為數不小的修繕清潔費 建議投標前應理性評估 心得 台積電治百病 預售屋 成屋 法拍屋都搶翻 下次要搶什麼物件呢

連結 內文 台積電帶旺高雄房市 成為全台目光 代銷不侷限在接待中心賣屋 有代銷業者銷售位於 美術館特區預售案 連續2週上竹科為工程師做簡報介紹 第一週一口氣就買下10幾戶 第二週再買下67戶 總價約1至2千萬之間的2或3房型 買氣熱絡 業者表示 工程師很忙 透過網站了解建案 聽完簡報後就線上支付簽約金 等到有空才會下高雄基地看看 或請同事代看 這些人經歷過



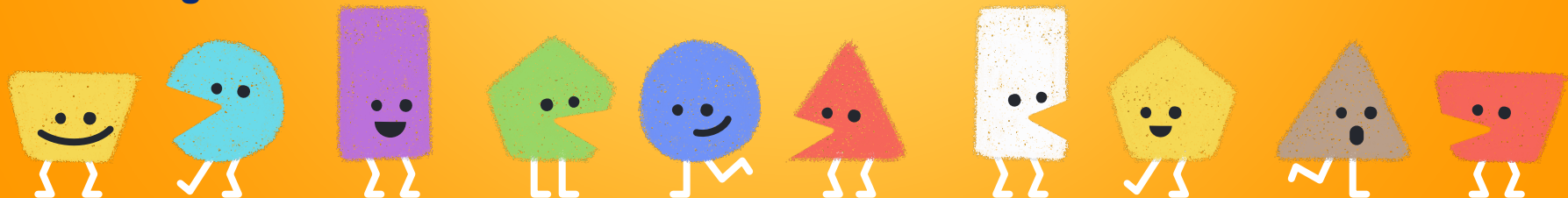


# 資料分析結果的發現與討論

透過相同的資料集，在不同ptt版別文字雲中可以看出，在不同的社群討論版中討論重點、熱度與情緒趨勢，即使是針對相同主題的分析，在不同類別的社群討論中，所著重關注的焦點亦可能有所不同。

本次分析主題原本預期會看見正面或負面情緒的明顯差距，正面(台積電帶來更多科技工作機會、高薪、人才流)，或者負面(帶來汙染、高房價)，但是我們在ptt八卦版裡的情緒分析結果發現，正反面情緒文章的篇數，並沒有顯著的差距。因此推斷有關本分析主題的正、負面評價略可稱衡平。

雖然分析主題和資料集皆相同，但是在不同社群版別進行分析時，關鍵字的提取、停用字的選擇，仍需依據不同社群討論主題的特性與趨勢，再透過觀察n-gram與詞頻計算之結果而隨之變動。







THANK YOU

