



探討網路民眾對於民眾黨的定位

政治所 碩一
生醫電資所 碩二
生醫電資所 碩二

R12322010 李安妮
R11945020 黃宣融
R11945022 葉冠均



Outline

- 目的
- 方法
- 研究流程
- **coding**
- 分析結果
- 結論



目的

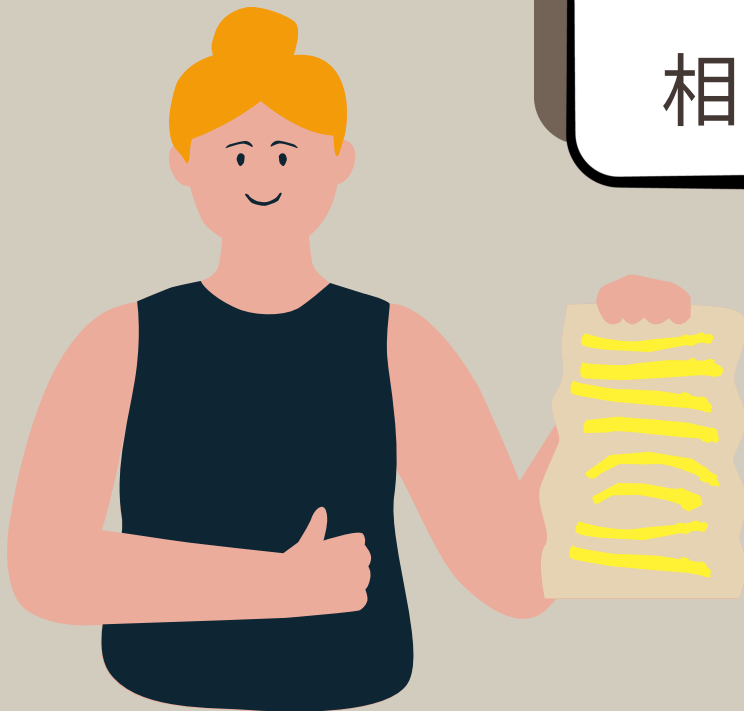
藉由網路民眾討論小黨政治相關的關鍵字之情緒來判斷民眾黨在選民心中的定位，以及網路生態與實際民調的關聯。

研究動機

過去自詡小黨的民眾黨，近年來在網路討論度高居不下是其他兩大政黨不具備的，因此聲勢不斷壯大，這樣的趨勢是否會造成選民對於其小黨定位迷失且民眾黨生量是否有效轉化為選票，過去吸收中立選民的優勢消失進而流失中立選票，甚至造成其追求成為關鍵少數的路程更加艱困。

方法

對**PTT**政黑板(因流量最大)的關鍵字留言進行情緒分析後，比對當月各黨民調結果是否有相關性，並探討其原因。



研究流程

爬蟲

篩選關鍵字

情緒分析

繪圖

每個月期間隨機挑選100頁：

- 7/27~8/24
- 8/25~9/26
- 9/27~10/24
- 10/25~11/26
- 11/27~12/12

近一個月(1000頁 data)



coding

爬蟲

```
from bs4 import BeautifulSoup
import pprint as pprint
import urllib.parse
import time
import requests

index = str(input('想抓取哪個ptt看板? (ex: Movie版請輸入 https://www.ptt.cc/bbs/movie/index.html): \n'))
pages = eval(input('想抓取幾頁呢? ex: 5: '))

not_exist = BeautifulSoup('<a>(本文已被刪除)</a>', 'lxml').a

def get_comments_on_article(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.text, 'lxml')
    comments = []

    main_content = soup.find('div', id='main-content')
    if main_content:
        push_elements = main_content.find_all('div', class_='push')
        for push in push_elements:
            push_info = list(push.stripped_strings)
            if len(push_info) >= 4:
                comment = {
                    'push_tag': push_info[0],
                    'push_user': push_info[1],
                    'push_content': push_info[2],
                    'push_time': push_info[3]
                }
                comments.append(comment)

    return comments
```

coding

爬蟲

```
def get_articles_and_comments(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.text, 'lxml')
    articles = []

    for i in soup.find_all('div', 'r-ent'):
        meta = i.find('div', 'title').find('a') or not_exist
        article_url = urllib.parse.urljoin(url, meta.get('href'))
        articles.append({
            'title': meta.getText().strip(),
            'url': article_url,
        })

    return articles

def get_pages(num):
    page_url = index
    all_articles = []

    for j in range(num):
        response = requests.get(page_url)
        articles = get_articles_and_comments(page_url)
        for article in articles:
            article_comments = get_comments_on_article(article['url'])
            article['comments'] = article_comments
            all_articles.append(article)
        page_url = urllib.parse.urljoin(index, BeautifulSoup(response.text, 'lxml').find('div', 'btn-group-paging').find_all('a', 'btn')[1].get('href'))

    return all_articles
```

coding

爬蟲

```
data = get_pages(pages)

for article in data:
    pprint.pprint(article)

csv_or_not = input('輸入 y 以匯出成csv檔，輸入其他結束程式：')

if csv_or_not == 'y':
    board = index.split('/')[2]
    csv = open('./ptt_%s版_前%d頁_評論.csv' % (board, pages), 'a+', encoding='utf-8')
    csv.write('標題, 評論內容, 評論時間, 評論者, \n')
    for article in data:
        for comment in article['comments']:
            csv.write(article['title'] + ',' + comment['push_content'] + ',' + comment['push_time'] + ',' + comment['push_user'] + ',\n')
    csv.close()
    print('CSV檔案已儲存在您的資料夾中。')
else:
    quit()
```


coding

情緒分析SnowNLP

```
! pip install snownlp
from snownlp import SnowNLP
```

#讀取資料

```
import pandas as pd
df = pd.read_csv('ptt_HatePolitics版_前100頁_12月.csv')
title = df.iloc[:, [0, 1]]
title.rename(columns={'標題': 'target', '評論內容': 'comment'}, inplace=True)
print(title)
```

	target	comment
0	[轉錄] 且看綠官們如何面對禮義廉恥	: 中國的恥在於台灣民主
1	[轉錄] 且看綠官們如何面對禮義廉恥	: 壞人的恥 在於被警察抓
2	[轉錄] 且看綠官們如何面對禮義廉恥	: 無恥之人 慣於以立場定恥與是非
3	[轉錄] 且看綠官們如何面對禮義廉恥	: 讀了廉恥就有廉恥?不讀廉恥就不知廉恥?
4	[轉錄] 且看綠官們如何面對禮義廉恥	: 超買不是走私!
...
85904	Re: [新聞] 高虹安告周玉蔻影射「與有婦之夫交往」不	: 看新聞報導的論述, 好像是有重疊到是之後
85905	Re: [新聞] 高虹安告周玉蔻影射「與有婦之夫交往」不	: 才離婚
85906	Re: [新聞] 高虹安告周玉蔻影射「與有婦之夫交往」不	: 沒離婚前 就是小三阿
85907	Re: [新聞] 高虹安告周玉蔻影射「與有婦之夫交往」不	: 李 就不敢講何時離婚啊 公布就一刀斃命w
85908	[討論] 阿北是否在用身體實踐科學	: 請正名柯學

[85909 rows x 2 columns]

coding

情緒分析SnowNLP

```
#民眾黨
#篩選標題中的小黨關鍵字
#篩target
keywords = ['柯文哲', '民眾黨', '白色力量', 'kp', '柯P', '阿北', '柯', '白', '柯粉', 'KP', '黃國昌', '國蕙', '蕙']
filtered_df = title[title['target'].str.contains('|'.join(keywords), case=False)]
print(filtered_df)
```

	target	comment
24	[新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨	: 區域別做夢啦 整天打雞血
25	[新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨	: 蔡壁如.....
26	[新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨	: 新竹民眾黨又沒提名
27	[新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨	: 柯妹哪有機會啦？
28	[新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨	: 國民黨都不敢選了才給你
...
85878	Re: [討論] 快訊！柯粉男神黃國昌下午要打賴老家！	: 黨當可以監督啊!!!
85879	Re: [討論] 快訊！柯粉男神黃國昌下午要打賴老家！	: 日本民眾也會監督在野黨啊。
85880	Re: [討論] 快訊！柯粉男神黃國昌下午要打賴老家！	: 南投國民黨是執政黨
85881	Re: [討論] 快訊！柯粉男神黃國昌下午要打賴老家！	: 綠共急了
85908	[討論] 阿北是否在用身體實踐科學	: 請正名柯學

[33165 rows x 2 columns]

```
#篩comment
keywords1 = ['小黨', '席位', '席次', '國會', '15%', '第三政黨', '立委', '少數', '關鍵', '15趴', '10趴', '11趴', '政黨票', '第']
filtered_df1 = filtered_df[filtered_df['comment'].str.contains('|'.join(keywords1), case=False)]
print(filtered_df1)
```

	target	comment
39	[新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨	: 笑死 說自己墨綠 然後立委反綠
65	[新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨	: 其實有稍微跟過那些立委候選人發言的就知
73	[新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨	: 統 政黨票雙崩耶 同時也代表TMD內參完全
97	[新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨	: 不分區也就34席 你拿15席粗估政黨票也要拿
107	[新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨	: 柯黨區域立委候選人魅力不足，名氣不夠，
...
85327	[討論] 柯現在是不是要祈禱兩件事	: 民眾黨關不關鍵，還是沒有他的票數重要
85512	[討論] 快訊！柯粉男神黃國昌下午要打賴老家！	: 萬里也是黃國昌當立委時的選區，上
85547	[討論] 快訊！柯粉男神黃國昌下午要打賴老家！	: 不分區立委不好好提自己政見是在？
85763	[新聞] 柯文哲如何說服入黨？黃國昌秀2人會議紀	: 第三點就是服貿 直說就好 不用堆砌文字
85839	[討論] 李登輝：柯文哲實在說起來，思想有問題	: 萬里也是黃國昌當立委時的選區，上

[414 rows x 2 columns]

coding

情緒分析SnowNLP

#篩選評論欄

```
key_comment = filtered_df1.iloc[:, 1]  
print(key_comment)
```

```
39          : 笑死 說自己墨綠 然後立委反綠  
65          : 其實有稍微跟過那些立委候選人發言的就知  
73          : 統 政黨票雙崩耶 同時也代表TMD內參完全  
97          : 不分區也就34席 你拿15席粗估政黨票也要拿  
107         : 柯黨區域立委候選人魅力不足, 名氣不夠,  
          ...  
85327       : 民眾黨關不關鍵, 還是沒有他的票數重要  
85512       : 萬里也是黃國昌當立委時的選區, 上  
85547       : 不分區立委不好好提自己政見是在?  
85763       : 第三點就是服貿 直說就好 不用堆砌文字  
85839       : 萬里也是黃國昌當立委時的選區, 上  
Name: comment, Length: 414, dtype: object
```

```
comment = str(key_comment)  
s = SnowNLP(comment)  
#所有句子的情緒分析  
print(s.sentiments)
```

1.3620768213318257e-07



數值：**0~1**
越接近**1**表示正面情緒，
越接近**0**表示負面情緒

民眾黨&小黨關鍵字

keywords = ['柯文哲','民眾黨','白色力量','kp','柯P','阿北','柯','白','柯粉','KP','黃國昌','國蔥','蔥']

keywords1 = ['小黨','席位','席次','國會','15%','第三政黨','立委','少數','關鍵','15趴','10趴','11趴','政黨票','第三','三大']

國民黨(KMT)關鍵字

blue_key = ['侯友宜','國民黨','侯侯做代誌','侯','警察','侯侯','藍','在野','侯粉','親中','馬英九','朱立倫','新北','趙少康','侯康','藍營','KMT']

民進黨(DPP)關鍵字

green_key = ['賴清德','民進黨','美德配','ㄌㄨˊㄨˊ','賴','賴蕭','蕭美琴','執政','綠','反中','蔡英文','ㄘㄨˊㄨˊ','蔡','綠營','DPP']

民眾黨(TPP)關鍵字

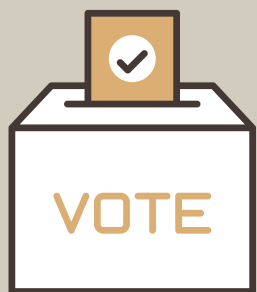
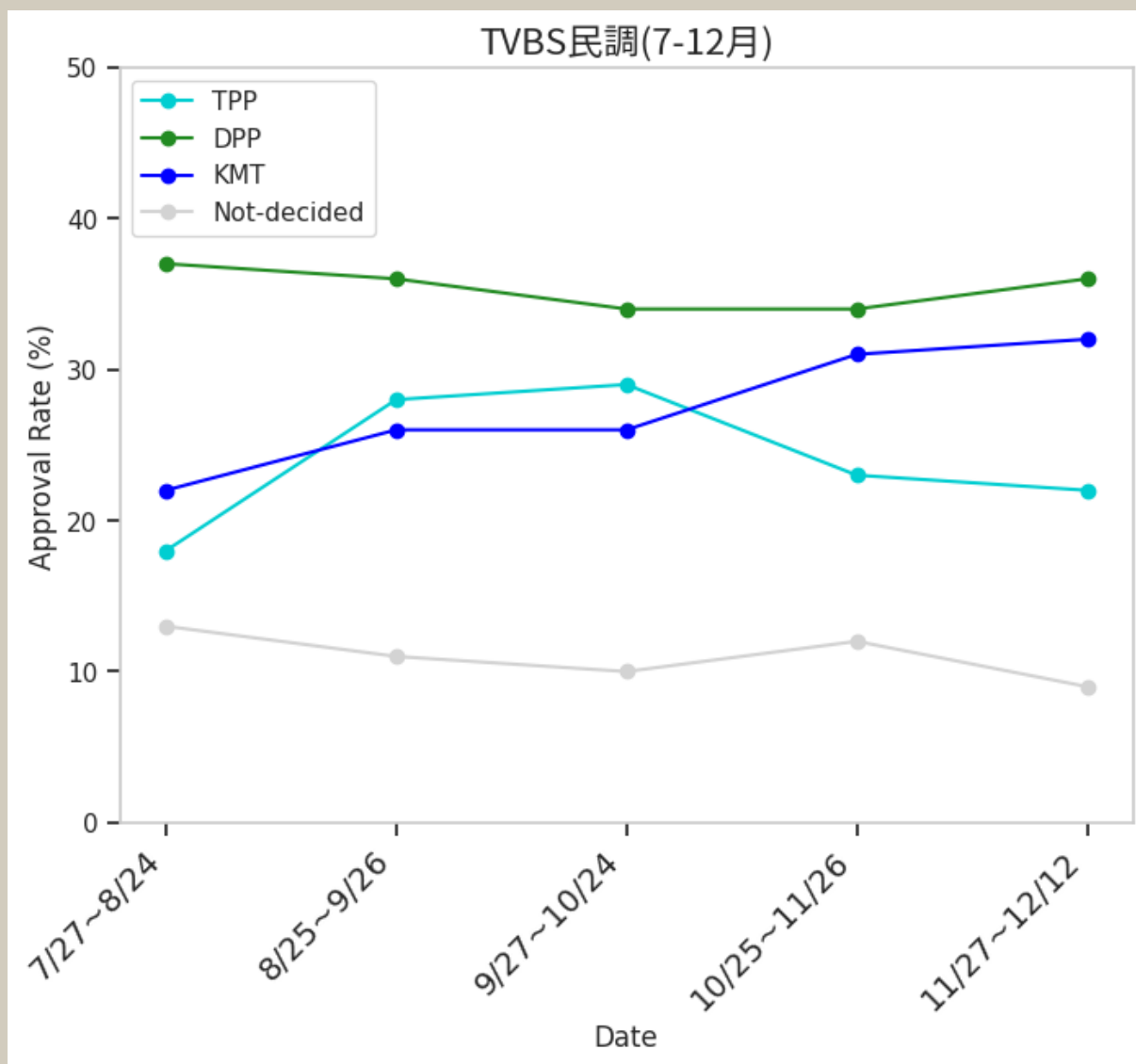
keywords = ['柯文哲','民眾黨','白色力量','kp','柯P','阿北','柯','白','柯粉','KP','黃國昌','國蔥','蔥','TPP']

```
#長條圖
#未篩選鄉民對於選舉的情緒分析
sns.set(style="whitegrid")
plt.figure(figsize=(8, 6))
sns.barplot(x='Date', y='Score of sentiment', data= score_df,hatch='//',color='peru')
plt.title('整體鄉民對於選舉之情緒分析', fontproperties='Taipei Sans TC Beta', fontsize=16)
plt.xlabel('Date', fontsize=16)
plt.ylabel('Score of sentiment (log scale)', fontsize=16)
plt.xticks(rotation=45, ha='right', fontsize=14)
plt.yticks(fontsize=14)
bar_width = 0.2
plt.ylim(0, 10)
plt.show()
```

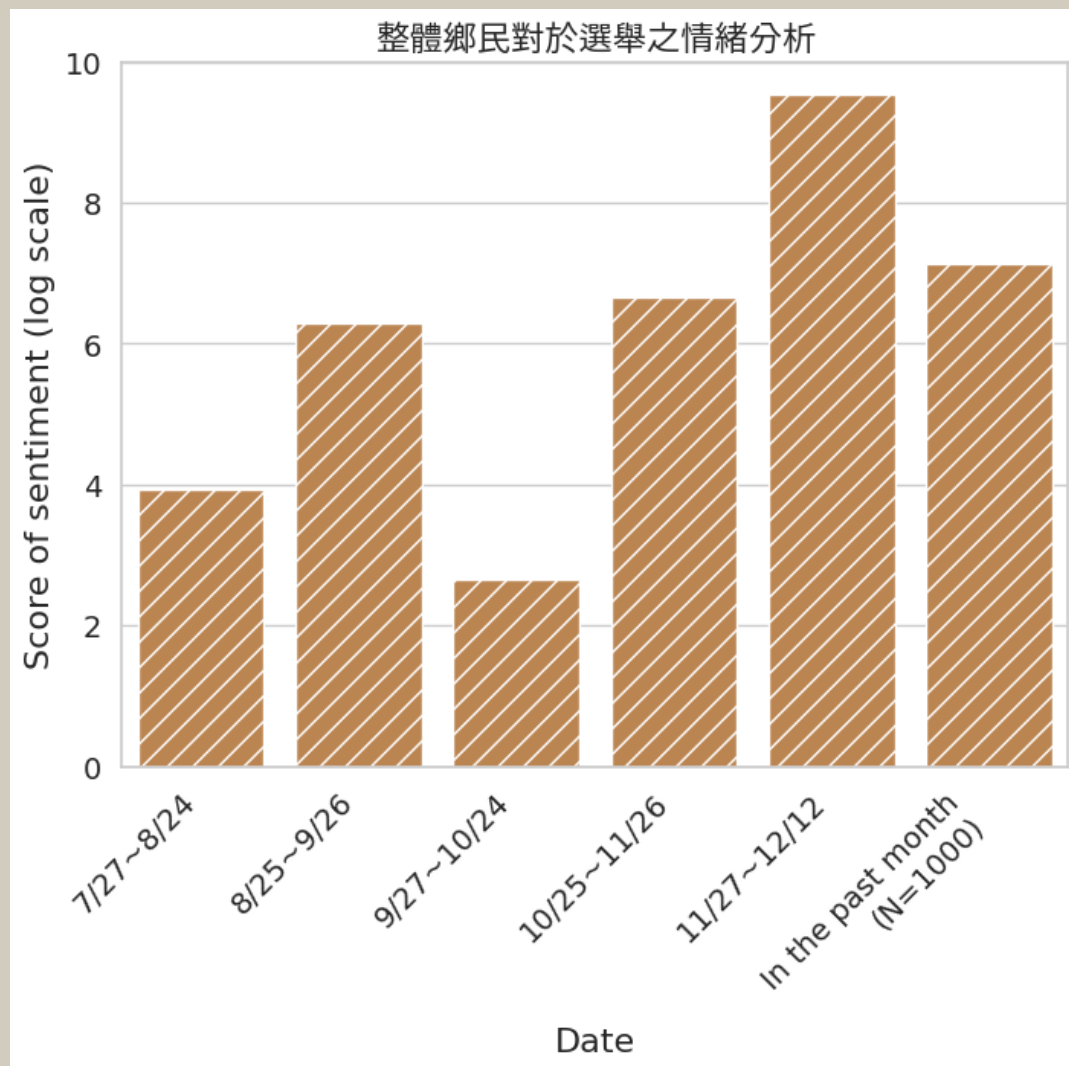
X軸為不同月份

Y軸以log scale的方式作呈現，因此高度越高代表其情緒越負面

TVBS 民調

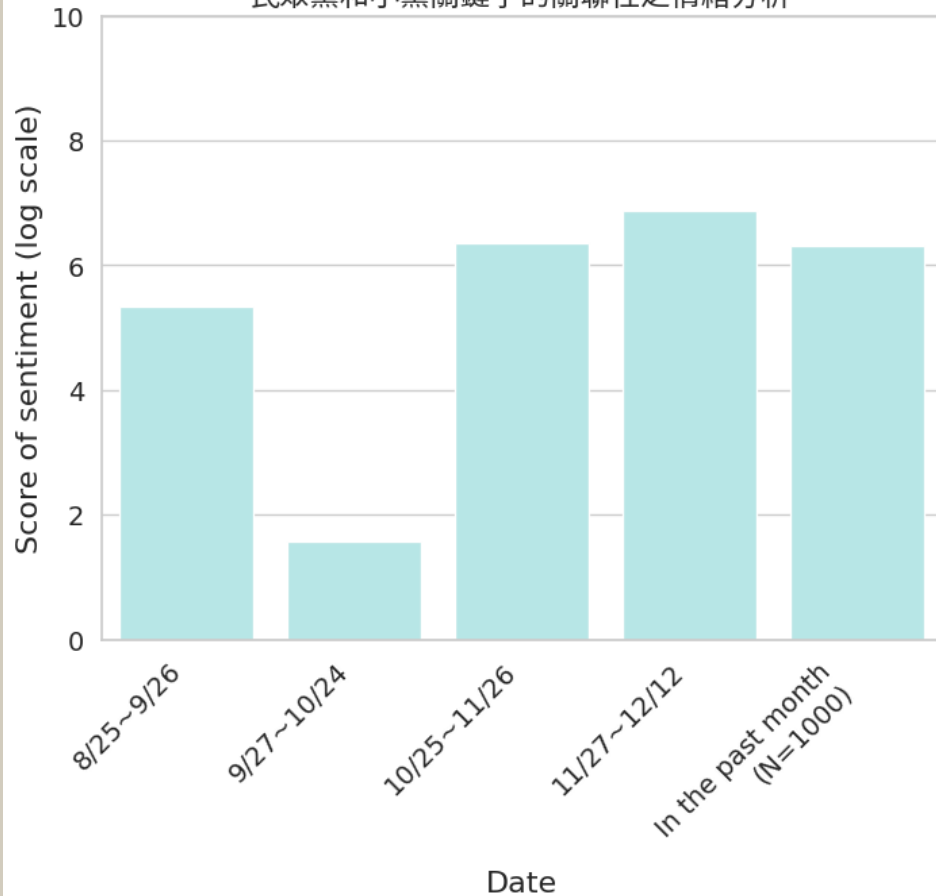


整體鄉民對於選舉之情緒分析

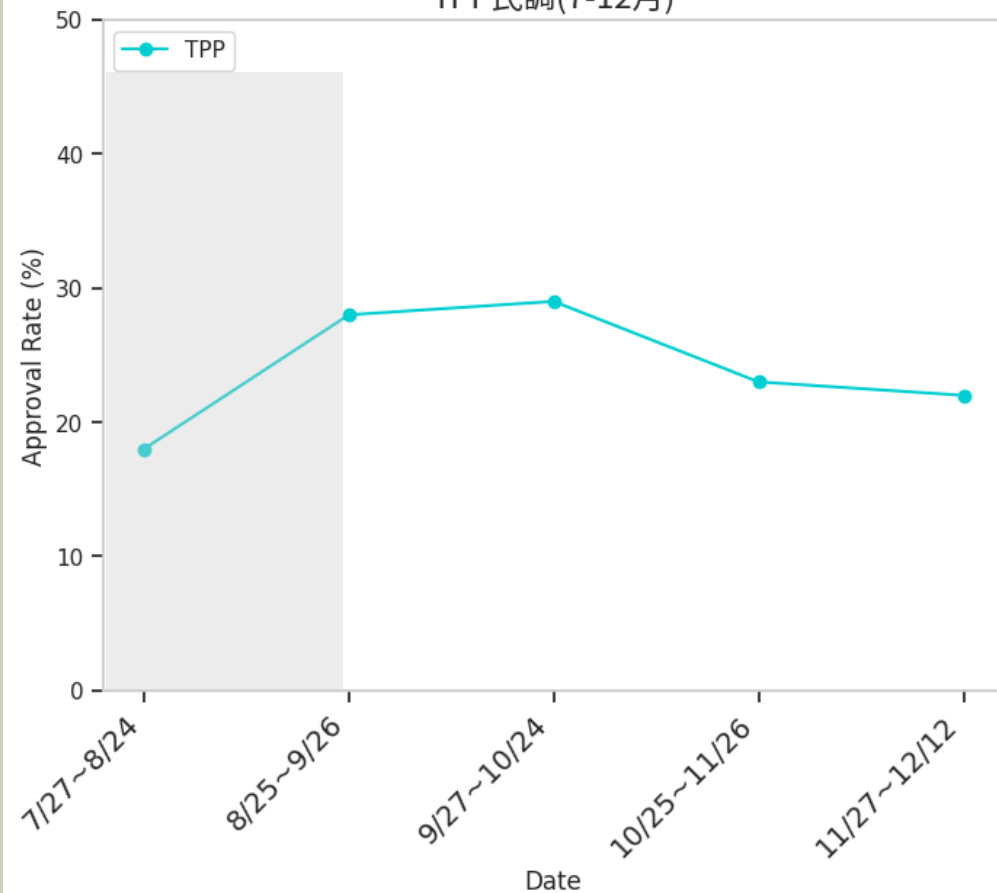


民眾黨和小黨關鍵字的關聯性 之情緒分析

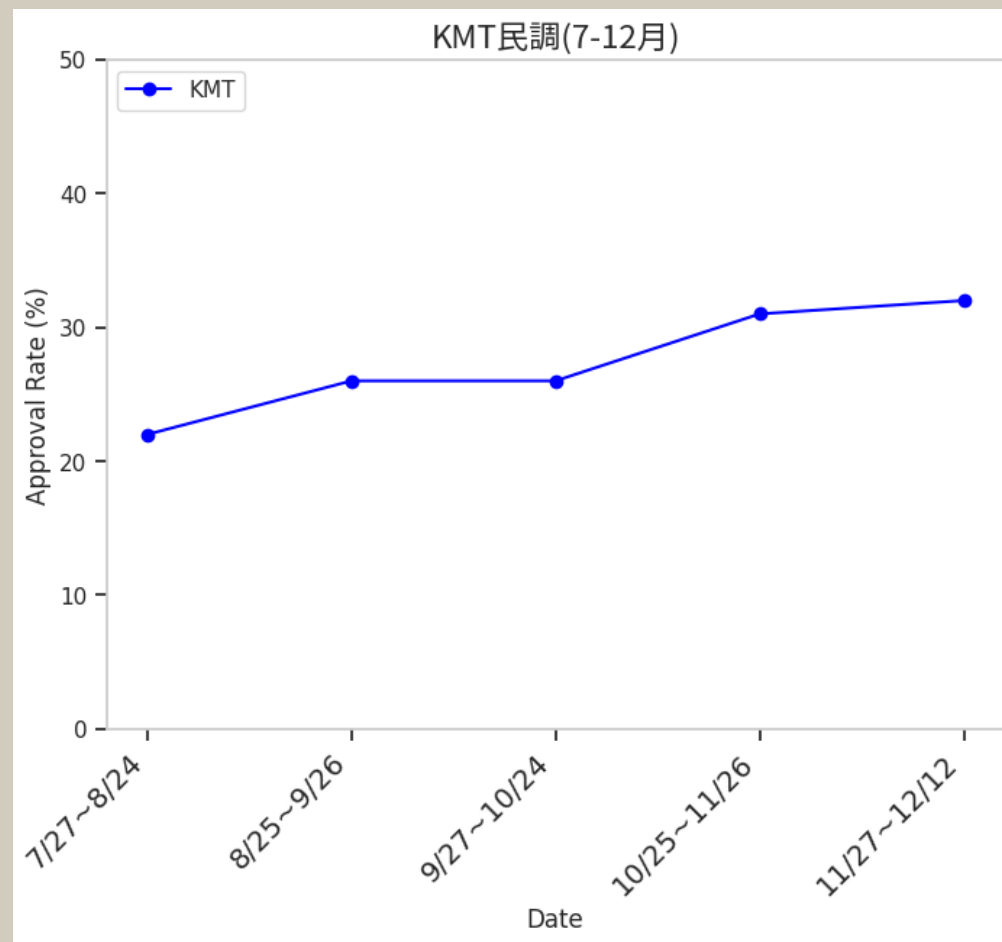
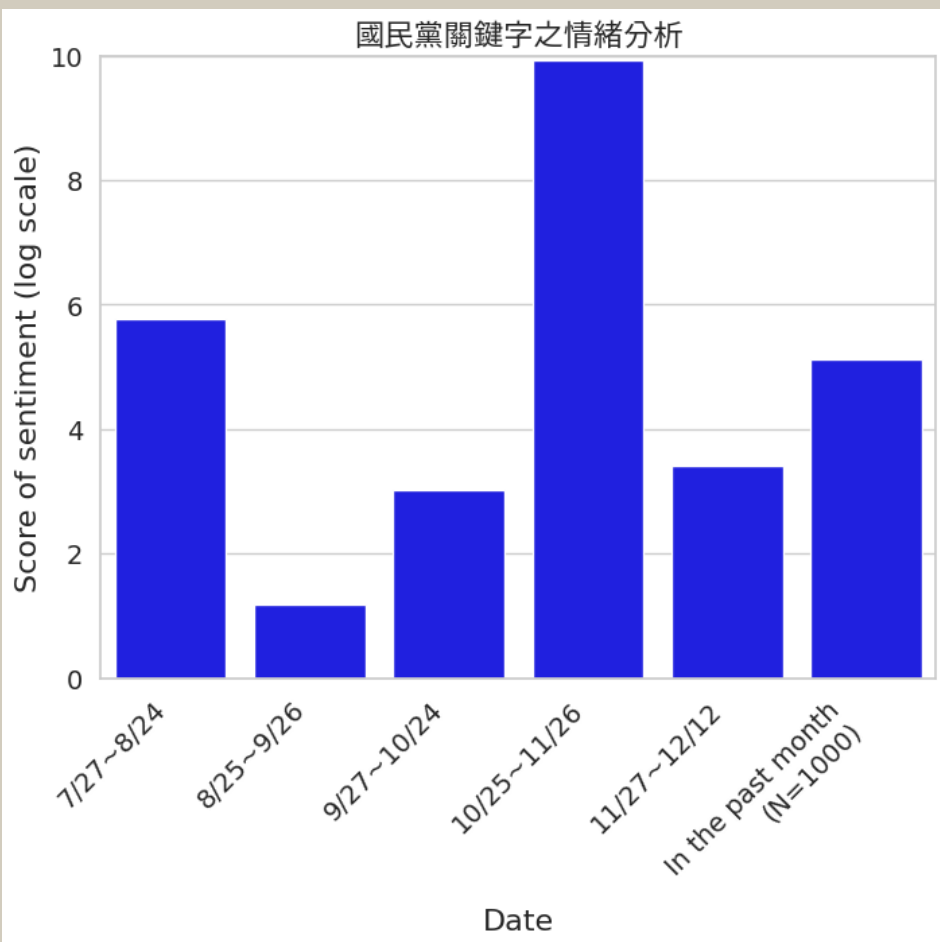
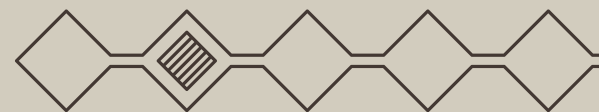
民眾黨和小黨關鍵字的關聯性之情緒分析

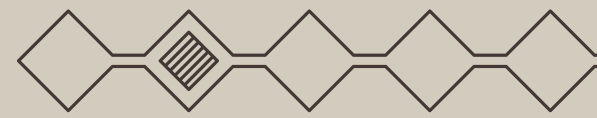


TPP民調(7-12月)



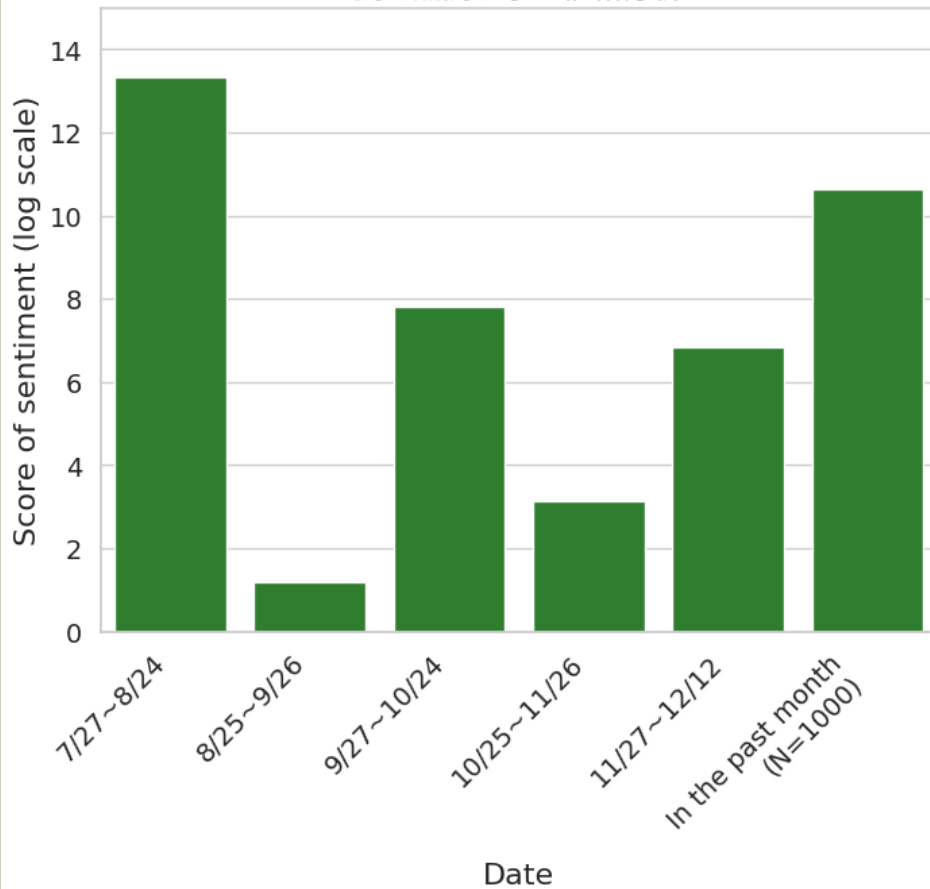
國民黨關鍵字的關聯性 之情緒分析



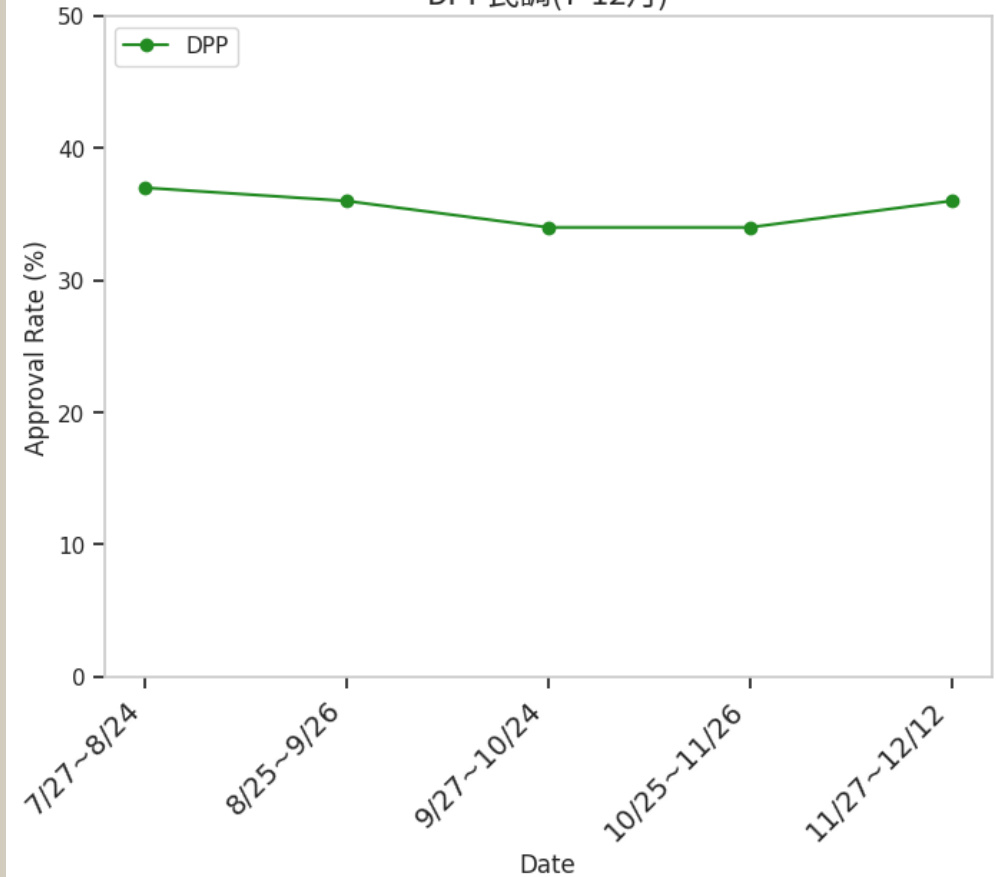


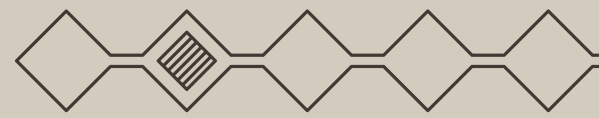
民進黨關鍵字的關聯性 之情緒分析

民進黨關鍵字之情緒分析

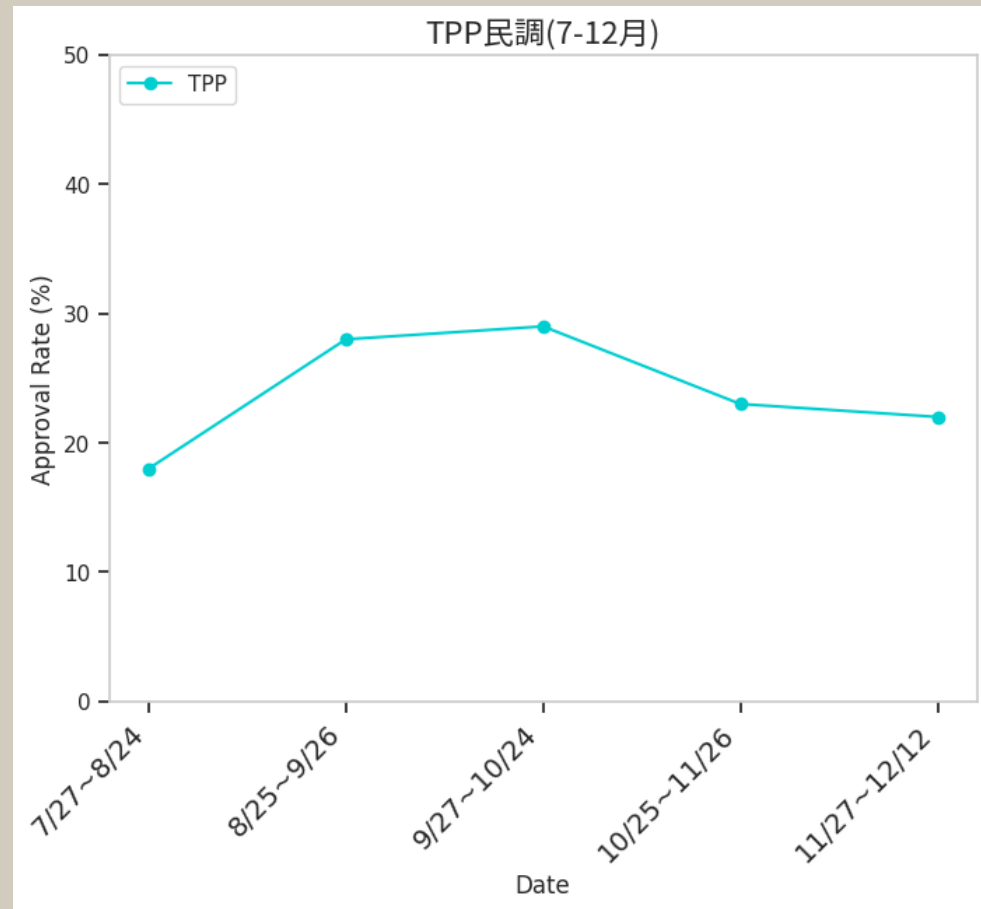
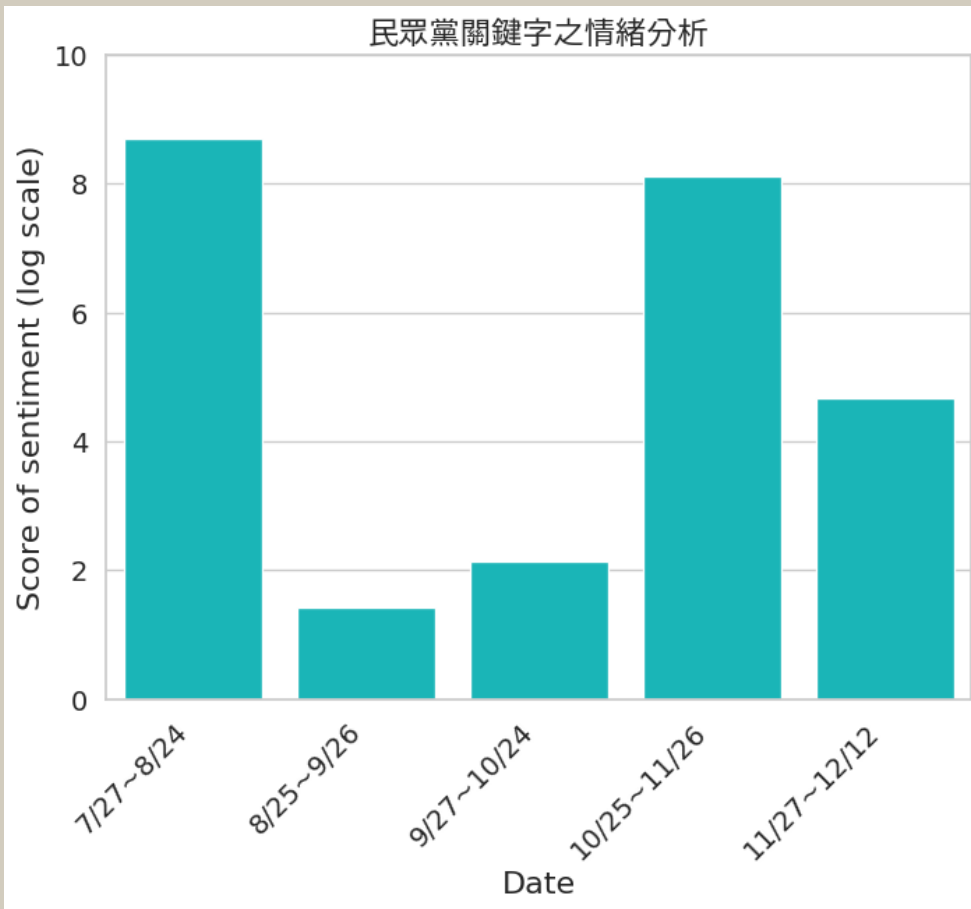


DPP民調(7-12月)





民眾黨關鍵字的關聯性 之情緒分析



結論



假設

若網路輿論與民調結果有正相關，可證明民眾黨在選民心中仍是小黨，在選舉結果中獲利且可能成為關鍵少數；

若無正相關，可能民眾黨在此次選戰的定位在網路民眾心中是相對模糊的，對其選情較不利，降低成為關鍵少數之可能。

經過前面的分析可以得知下面幾點：

- 民眾黨與小黨之關聯性存在，但是並不顯著
- 藍綠兩大政黨網路情緒聲量與民調完全脫鉤
- 在特別事件發生時，情緒聲量會有明顯改變



甘蝦恁!



加分

