

從論壇分析台灣人對反送中的討論度

問題描述：台灣人超級健忘？

在我們的觀察之下，
大多數的台灣人對於話題的熱度往往隨著時間便愈趨平淡，
無論是之前的統一地溝油事件，
甚至是攸關台灣未來的國際大事也不例外，
所以我們決定來分析近期最大的事件
「香港反送中」在論壇上的討論度，
是否和我們觀察的一樣，會隨著時間而愈來愈乏人問津。



反送中水深火熱?
沒拿捏好朋友分寸?
高雄發大財?



解決關鍵：抵制媒體亂象

前陣子在凱道舉行的「反紅媒遊行」，
以及NCC開罰中天及東森巨額罰款，
都可以看出台灣的媒體亂象越來越嚴重，
也有很多人已經發現這樣的現象，
媒體可以是第四權，也能是帶風向的好工具，
身為一個受過教育的大學生，我們應該要具備
良好的「媒體識讀」能力，
觀賞播報正確、立場中立的新聞台。

解決關鍵：善用Google，不再依賴懶人包

在我們蒐集資料的時候，大量瀏覽了許多論壇上的文章，多數文章立場並不中立，尤其是「**懶人包**」，這類文章多灌輸個人的意識形態，立場偏頗，資料不完整，在這個網路發達的時代，雖然網路上的資料多的數不清，要找資料很容易但要查到完整且正確的資料卻也像大海撈針，因此很多人都會因為懶惰，而只看懶人包，就聽信單一立場的言論，這樣是不對的。

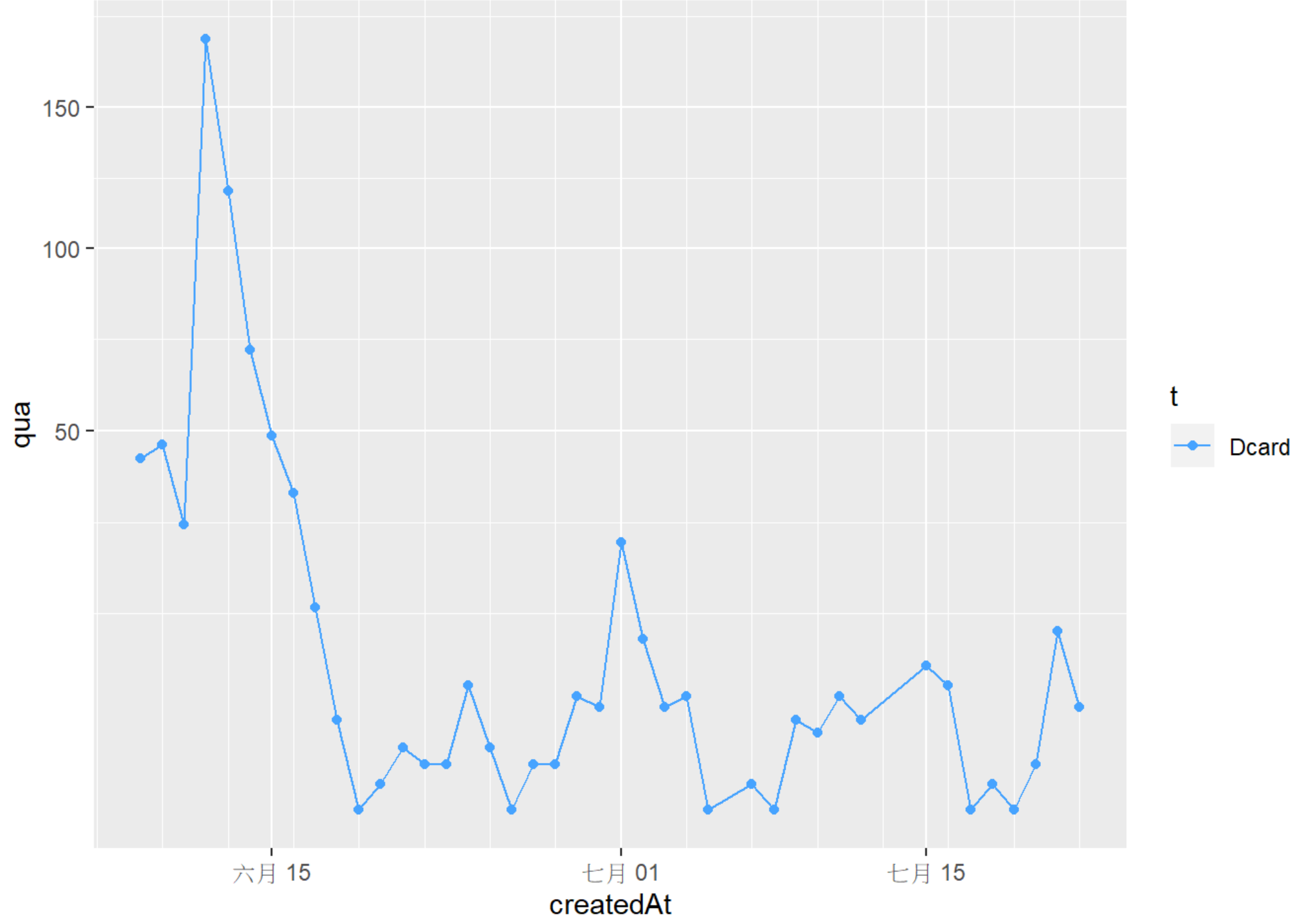
解決關鍵：不要盲從

這次的事件，讓很多平時不關注政治的人，也開始談論，但關注的人很多，但是實際上真正有持續關心的人又占多少？關心政治不能只是一天兩天的事，也不是在網路上發發文、看看懶人包就是關心，尤其是現在的政治情形，
影響的就是正值青春歲月的有為大學生的未來。

你還在政治冷感嗎？
快醒醒 這是你的未來！

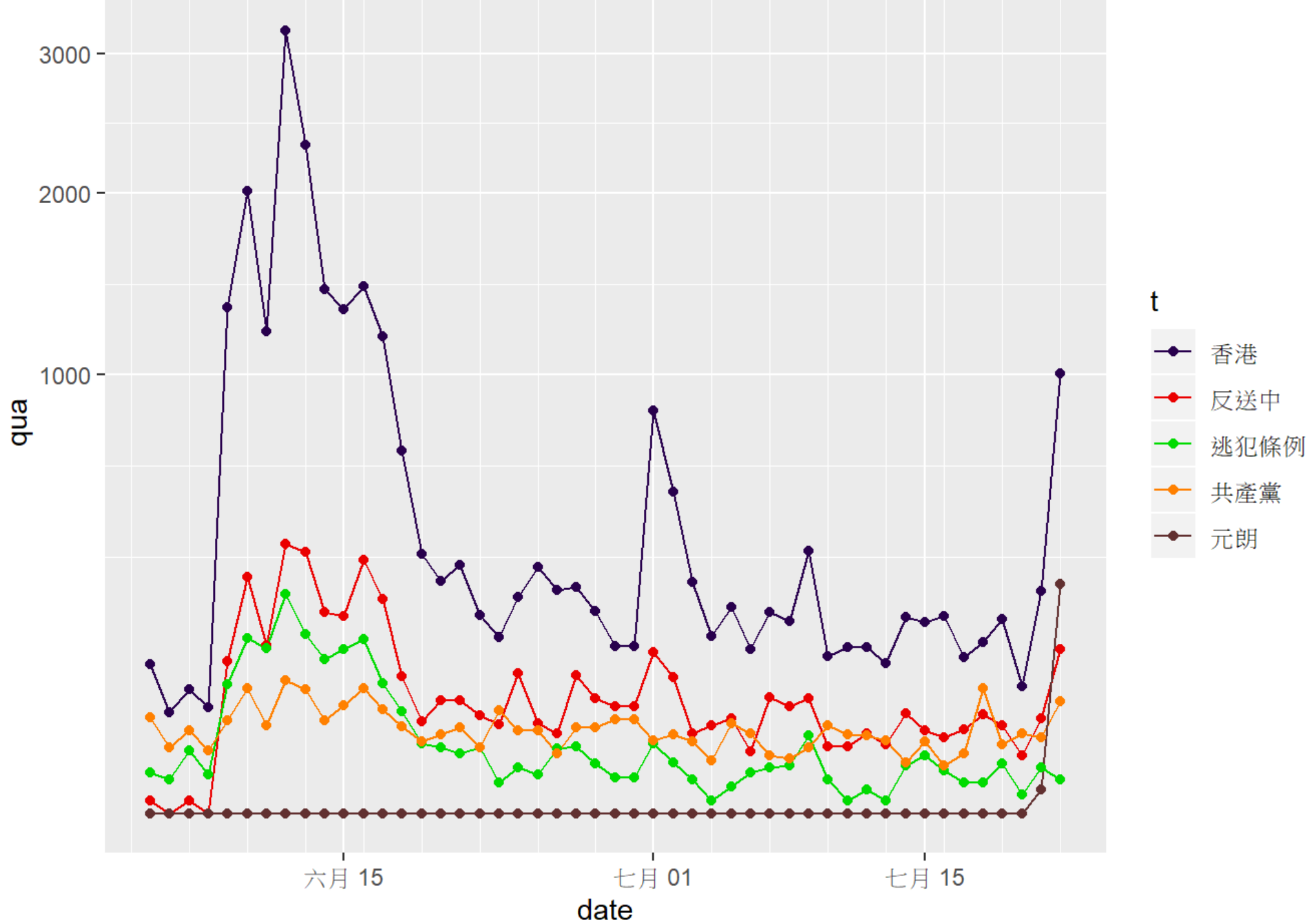
觀察重點

- 貼文數是否根據情勢維持一定水準，還是一日行情？
- 論壇在此議題是否為主要傳聲筒？

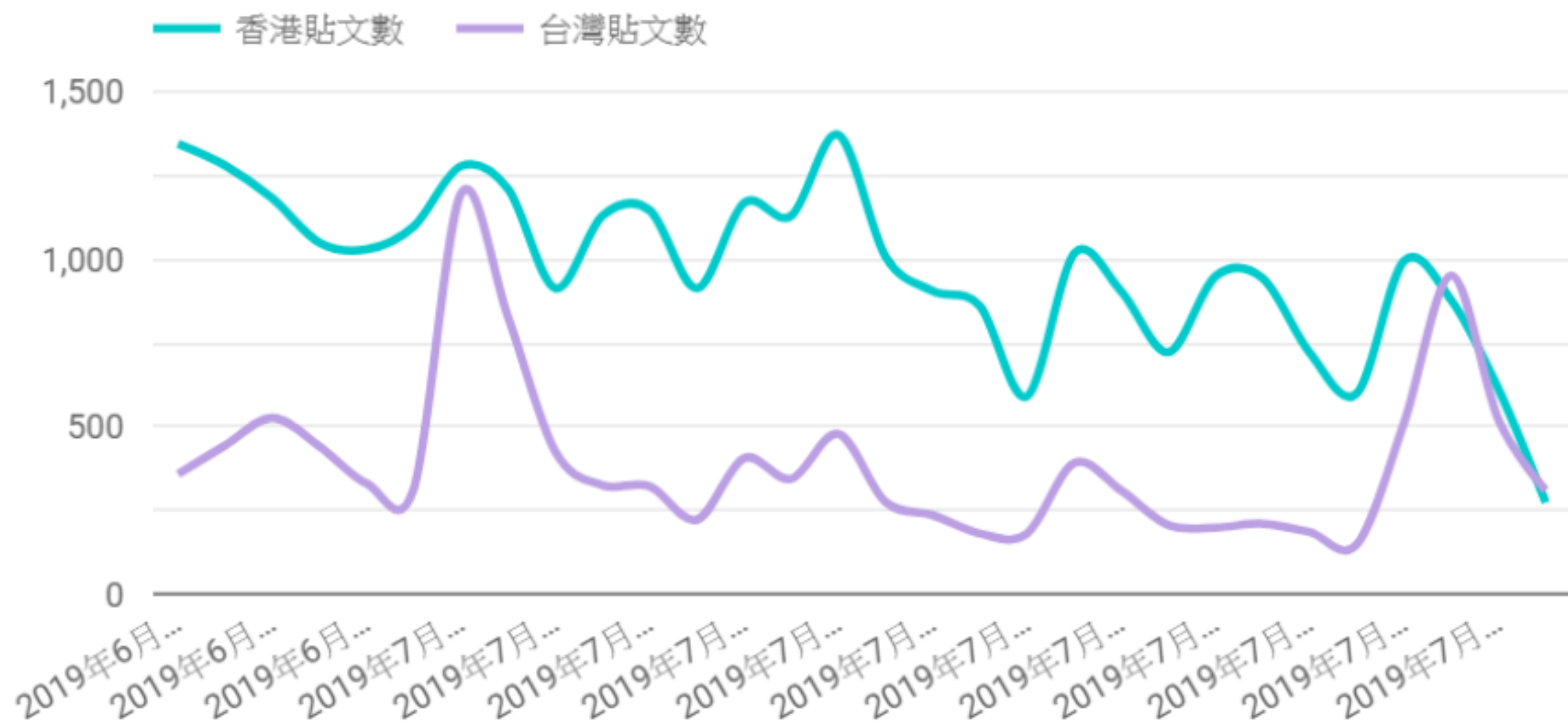


蔡英文
韓國瑜
國民黨
總統
台灣
政治
香港
民進黨
民意
大家
氣氣
關心
風向
黨工
民主
清場
垃圾車
夥伴
燙手山芋
民主
主人
記錄器
示威
果凍
民
前哨戰
智力測驗
全國機器聲明
過境
代表
消息
警司
生氣
上線
韓粉
風波
三毛
大選
方法
方車
錄影
光頭
示威者
畢業
速度
隊友
人物對話
冠軍
短期
票內
聲音
開票
中國
香蕉
信守
市長
升官
邏輯
卡車
參賽
指揮官
學歷
爸爸
樓主
暴力
形式
戰神
智力
宇宙
運轉
人
奇蹟
煎餅
智商測驗
拳手
產品
險情
噁心
選民
遊樂園
事情
全力支持
假裝
行車
移民
關心
風向
民意
氣氣
大家
記錄器
清場
垃圾車
夥伴
燙手山芋
民主
主人
示威
果凍
民
前哨戰
智力測驗
全國機器聲明
過境
代表
消息
警司
生氣
上線
韓粉
風波
三毛
大選
方法
方車
錄影
光頭
示威者
畢業
速度
隊友
人物對話
冠軍
短期
票內
聲音
開票
中國
香蕉
信守
市長
升官
邏輯
卡車
參賽
指揮官
學歷
爸爸
樓主
暴力
形式
戰神
智力
宇宙
運轉
人
奇蹟
煎餅
智商測驗
拳手
產品
險情
噁心
選民
遊樂園
事情
全力支持
假裝
行車
移民
關心
風向
黨工
民主
清場
垃圾車
夥伴
燙手山芋
民主
主人
示威
果凍
民

[illegible]



Excerpt



資料來源

- 透過爬蟲程式蒐集PTT、Dcard兩大論壇從反送中爆發至7/22的所有文章
- 上網找到的輔助資料：[反送中即時社群數據](#)

發文量統計：

香港總貼文量

29,214

↓ -42.2%

台灣總貼文量

11,773

↓ -53.9%

香港媒體總貼文量

12,364

↓ -30.8%

台灣媒體總貼文量

6,384

↓ -40.5%

互動量統計：

香港總互動量

1,256.3萬

↓ -44.1%

台灣總互動量

266.3萬

↓ -75.8%

香港媒體總互動量

816.3萬

↓ -46.5%

台灣媒體總互動量

137.3萬

↓ -72.8%

觀察重點

- 憤怒趨勢是否持續增長，代表這議題是否持續發燒
- 憤怒比率超過5%表示有潛力，**超過10% 表示很憤怒**。

憤怒溫度計

香港憤怒比率

18.9%

↓ -14.3%

台灣憤怒比率

10.9%

↑ 10.8%

— 香港憤怒比率

40%

20%

0%

2019年6月...
2019年6月...
2019年6月...
2019年7月...
2019年7月...
2019年7月...
2019年7月...
2019年7月...
2019年7月...
2019年7月...
2019年7月...
2019年7月...
2019年7月...
2019年7月...
2019年7月...



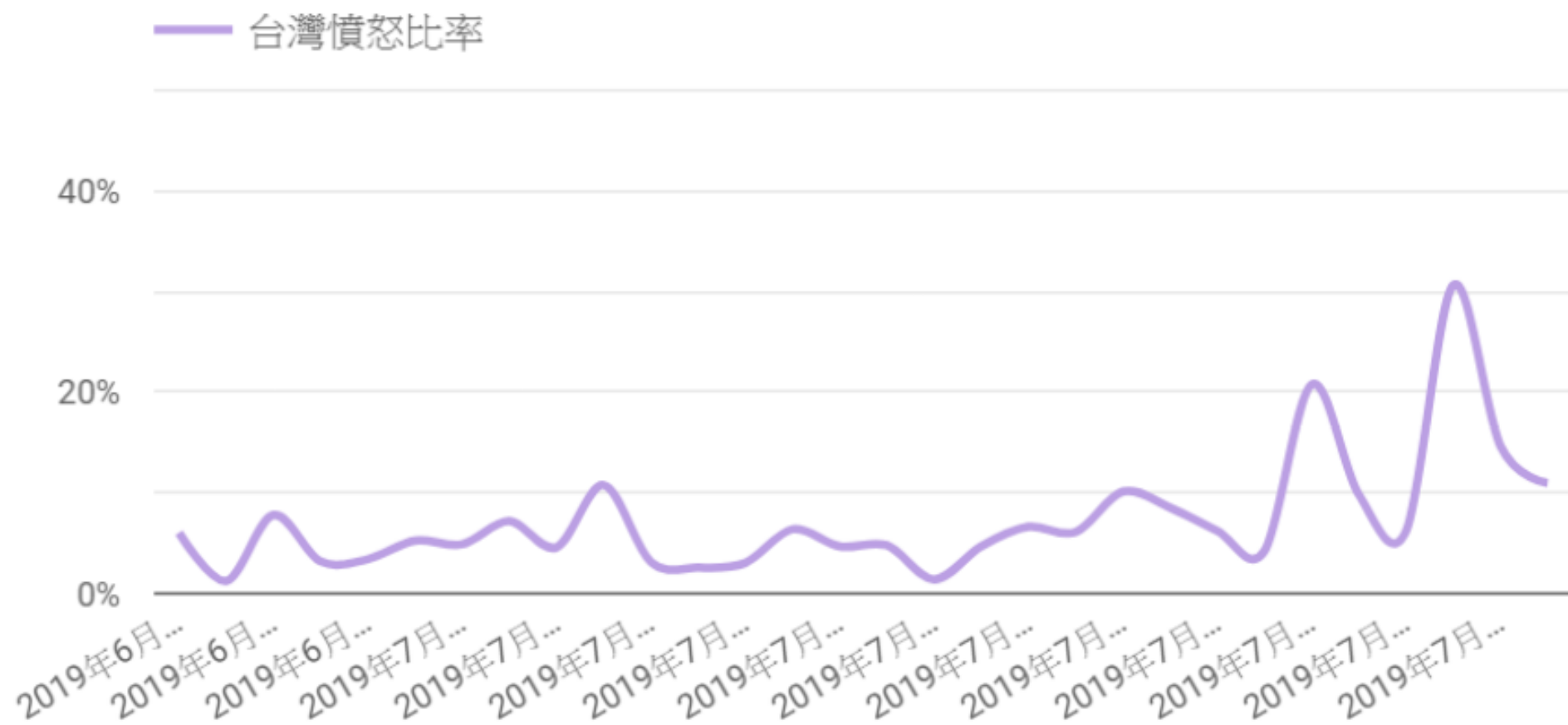


	圖	名稱	期間總發文數	反送中發文數	反送中文總影響力分數
1.		太紀元 epochtimes.com	2,866	381	19,210
2.		自由時報	2,870	122	19,785
3.		Yahoo!奇摩新聞	5,257	98	6,780
4.		蘋果新聞網	1,387	79	6,384
5.		三立新聞	5,501	73	10,870
6.		風傳媒	2,314	72	2,234
7.		民視新聞	1,641	61	2,343
8.		公視新聞網 PNN	501	57	12,822
9.		udn.com 聯合新聞網	2,445	53	5,443
10.		中央社新聞粉絲團	896	49	3,866
11.		東森新聞	3,839	33	12,888
12.		ETtoday新聞雲	4,419	25	10,587
13.		TVBS 新聞	2,804	21	3,496
14.		中時電子報	2,988	9	710
15.		中天新聞52家族	240	1	62

如何挑選進行分析屬性

名稱	批踢踢實業坊	Dcard
成立時間	1995/9/14	2011/12/16
註冊人數	150-200萬	約100萬
上線人數	100,000 (同時在線人數)	42,500 (每日參訪數)
年齡族群	25-45，上班族居多	18-24，大學生居多
討論內容	依各討論版類型不同，熱度最高的八卦版以政治、時事為主。	熱門議題、回應最踴躍以情感、課業相關討論為多。
Alexa排名	台灣27/世界1,061 (2017/11/1)	台灣14/世界494 (2017/11/1，此為網頁版數據)

階段性結果、分析的困難處與如何解決

processing article 125 / 2390 ...[1] "https://www.ptt.cc/bbs/Goss
processing article 126 / 2390 ...[1] "https://www.ptt.cc/bbs/Goss
processing article 127 / 2390 ...[1] "https://www.ptt.cc/bbs/Goss
processing article 128 / 2390 ...[1] "https://www.ptt.cc/bbs/Goss
processing article 129 / 2390 ...[1] "https://www.ptt.cc/bbs/Goss
processing article 130 / 2390 ...[1] "https://www.ptt.cc/bbs/Goss
processing article 131 / 2390 ...[1] "https://www.ptt.cc/bbs/Goss
processing article 132 / 2390 ...[1] "https://www.ptt.cc/bbs/Goss
processing article 133 / 2390 ...[1] "https://www.ptt.cc/bbs/Goss
processing article 134 / 2390 ...[1] "https://www.ptt.cc/bbs/Goss

11萬八卦版費雯文章
可能要爬8小時

```
jieba.R x sub1.R x sub2.R x sub3.R x sub4.R x sub5.R x sub6.R x
Source on Save
47 #tab <- data.frame(
48 # titles = NA,
49 # texts = NA,
50 # time.stamp = NA)
51 progress <- 0
52 abs.progress <- 0
53 from <- 10923
54 to <- 13345
55 total <- to - from + 1
56 for (i in from : to){
57   progress <- progress + 1
58   abs.progress <- abs.progress + 1
59   title <- title + 1
60   text <- text + 1
61   time <- time + 1
62   #print(paste("article", i, "progress", progress, "abs.progress", abs.progress, "title", title, "text", text, "time", time))
63 }
```

PTT時間戳章問題



PTT_gossiping_extradition_law

2019/7/25 上午 01:33

Microsoft Excel 逗點...

15,468 KB



























txtDate - 已儲存



























新細明體

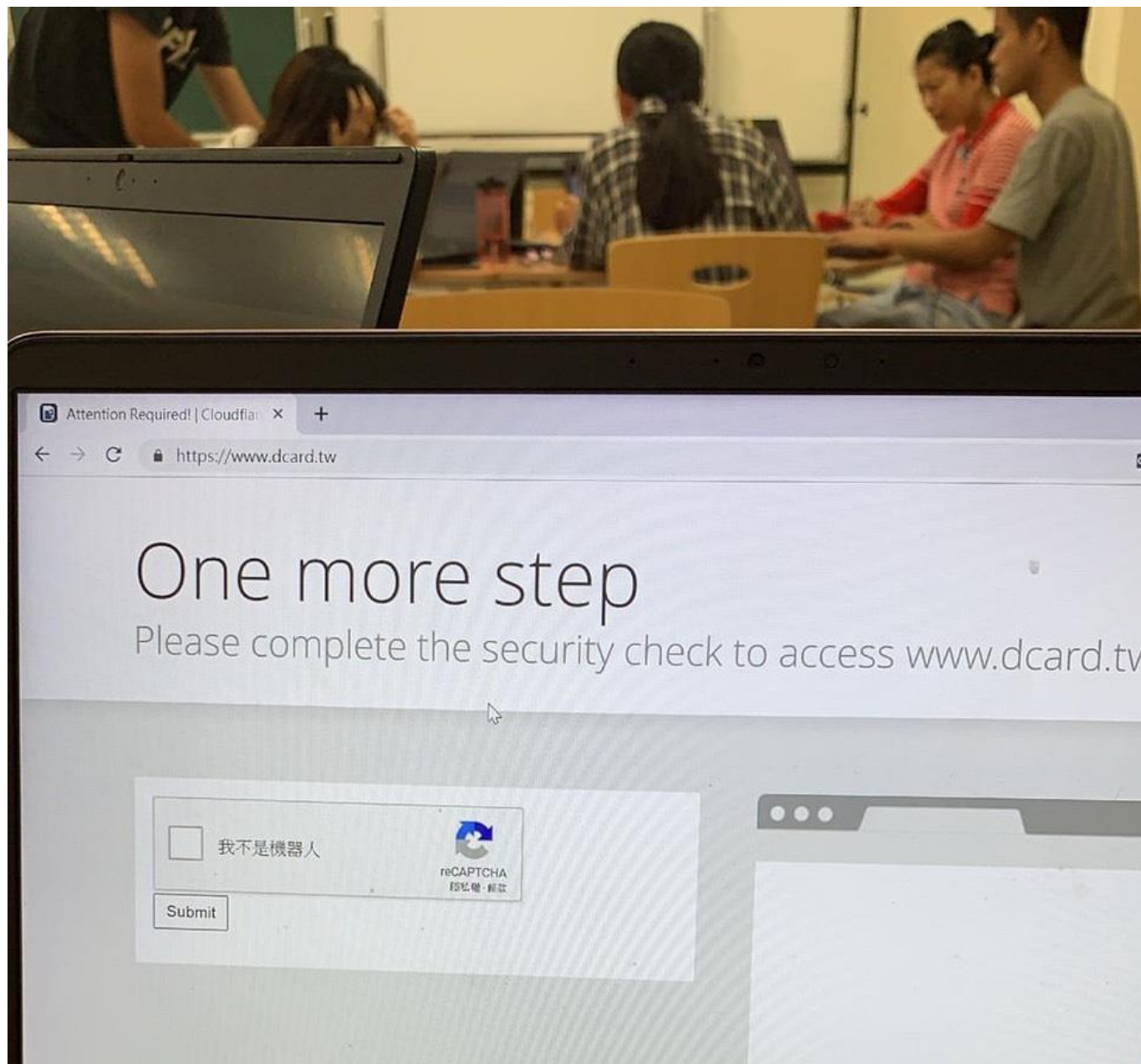
titleDate - 已儲存

txtDate - 已儲存

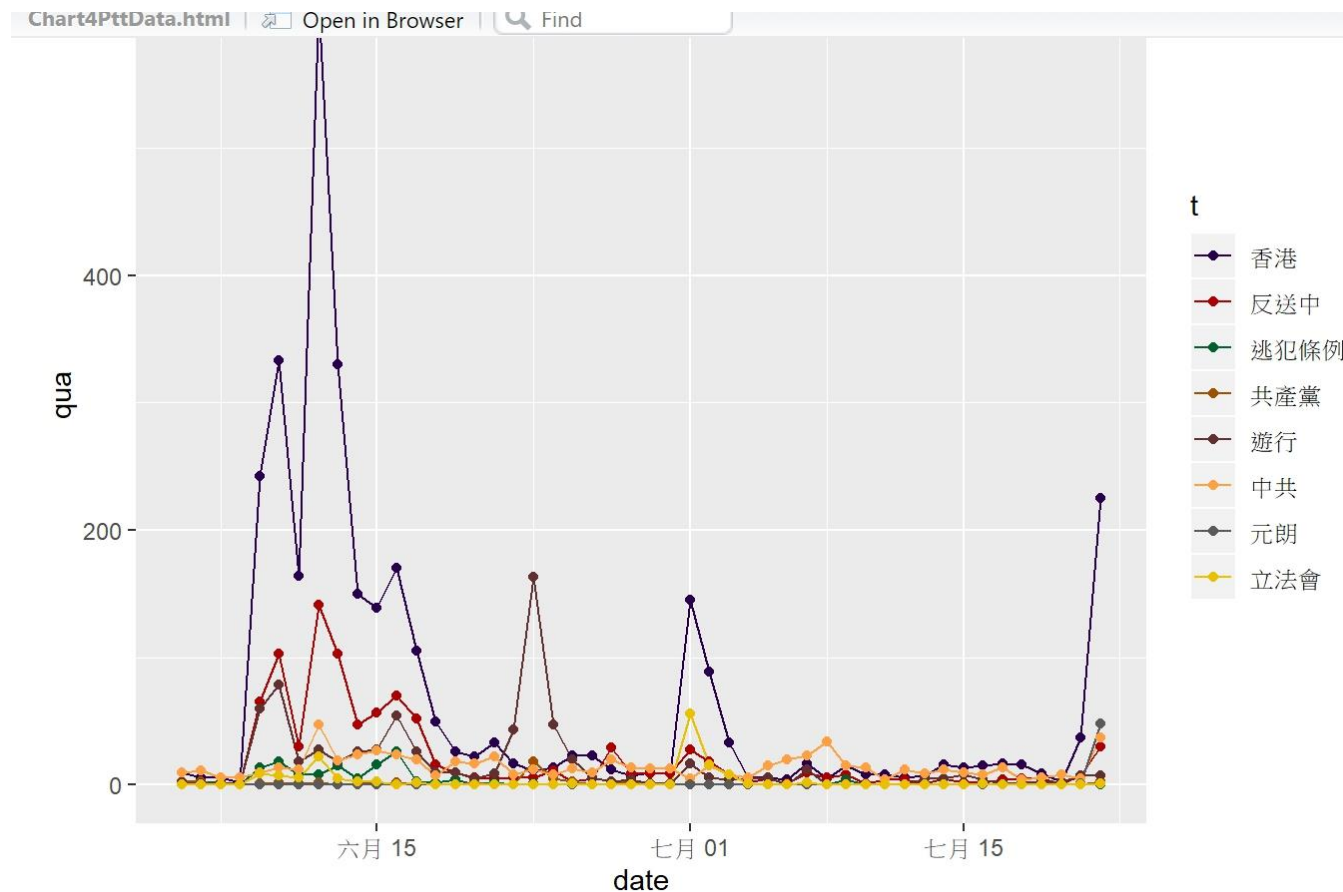


 data_Jul_1	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,460 KB
 data_Jul_2	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,636 KB
 data_Jul_3	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,476 KB
 data_Jul_4	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,727 KB
 data_Jul_5	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,825 KB
 data_Jul_6	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,359 KB
 data_Jul_7	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,396 KB
 data_Jul_8	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,361 KB
 data_Jul_9	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,526 KB
 data_Jul_10	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,393 KB
 data_Jul_11	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,453 KB
 data_Jul_12	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,478 KB
 data_Jul_13	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,307 KB
 data_Jul_14	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,290 KB
 data_Jul_15	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,460 KB
 data_Jul_16	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,452 KB
 data_Jul_17	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,566 KB
 data_Jul_18	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,638 KB
 data_Jul_19	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,603 KB
 data_Jul_20	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,361 KB
 data_Jul_21	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,373 KB
 data_Jul_22	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,743 KB
 data_Jun_5	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,622 KB
 data_Jun_6	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,486 KB
 data_Jun_7	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,412 KB
 data_Jun_8	2019/7/24 下午 01:50	Microsoft Excel 逗點...	1,348 KB

 DcardData119	2019/7/22 下午 02:45	Microsoft Excel 逗點...	46 KB
 DcardData139	2019/7/22 下午 02:46	Microsoft Excel 逗點...	46 KB
 DcardData159	2019/7/22 下午 02:46	Microsoft Excel 逗點...	47 KB
 DcardData179	2019/7/22 下午 02:47	Microsoft Excel 逗點...	46 KB
 DcardData199	2019/7/22 下午 02:47	Microsoft Excel 逗點...	45 KB
 DcardData219	2019/7/22 下午 02:48	Microsoft Excel 逗點...	46 KB
 DcardData239	2019/7/22 下午 02:48	Microsoft Excel 逗點...	48 KB
 DcardData259	2019/7/22 下午 02:48	Microsoft Excel 逗點...	48 KB
 DcardData279	2019/7/22 下午 02:49	Microsoft Excel 逗點...	49 KB
 DcardData299	2019/7/22 下午 02:49	Microsoft Excel 逗點...	48 KB
 DcardData319	2019/7/22 下午 02:50	Microsoft Excel 逗點...	47 KB
 DcardData339	2019/7/22 下午 02:50	Microsoft Excel 逗點...	48 KB
 DcardData359	2019/7/22 下午 02:51	Microsoft Excel 逗點...	47 KB
 DcardData379	2019/7/22 下午 02:51	Microsoft Excel 逗點...	46 KB
 DcardData399	2019/7/22 下午 02:52	Microsoft Excel 逗點...	47 KB
 DcardData419	2019/7/22 下午 02:52	Microsoft Excel 逗點...	48 KB
 DcardData439	2019/7/22 下午 02:53	Microsoft Excel 逗點...	47 KB
 DcardData459	2019/7/22 下午 02:53	Microsoft Excel 逗點...	49 KB
 DcardData479	2019/7/22 下午 02:53	Microsoft Excel 逗點...	49 KB
 DcardData499	2019/7/22 下午 02:54	Microsoft Excel 逗點...	47 KB
 DcardData519	2019/7/22 下午 02:54	Microsoft Excel 逗點...	49 KB
 DcardData539	2019/7/22 下午 02:55	Microsoft Excel 逗點...	50 KB
 DcardData559	2019/7/22 下午 02:55	Microsoft Excel 逗點...	50 KB
 DcardData579	2019/7/22 下午 02:56	Microsoft Excel 逗點...	50 KB
 DcardData599	2019/7/22 下午 02:56	Microsoft Excel 逗點...	52 KB
 DcardData619	2019/7/22 下午 02:57	Microsoft Excel 逗點...	50 KB

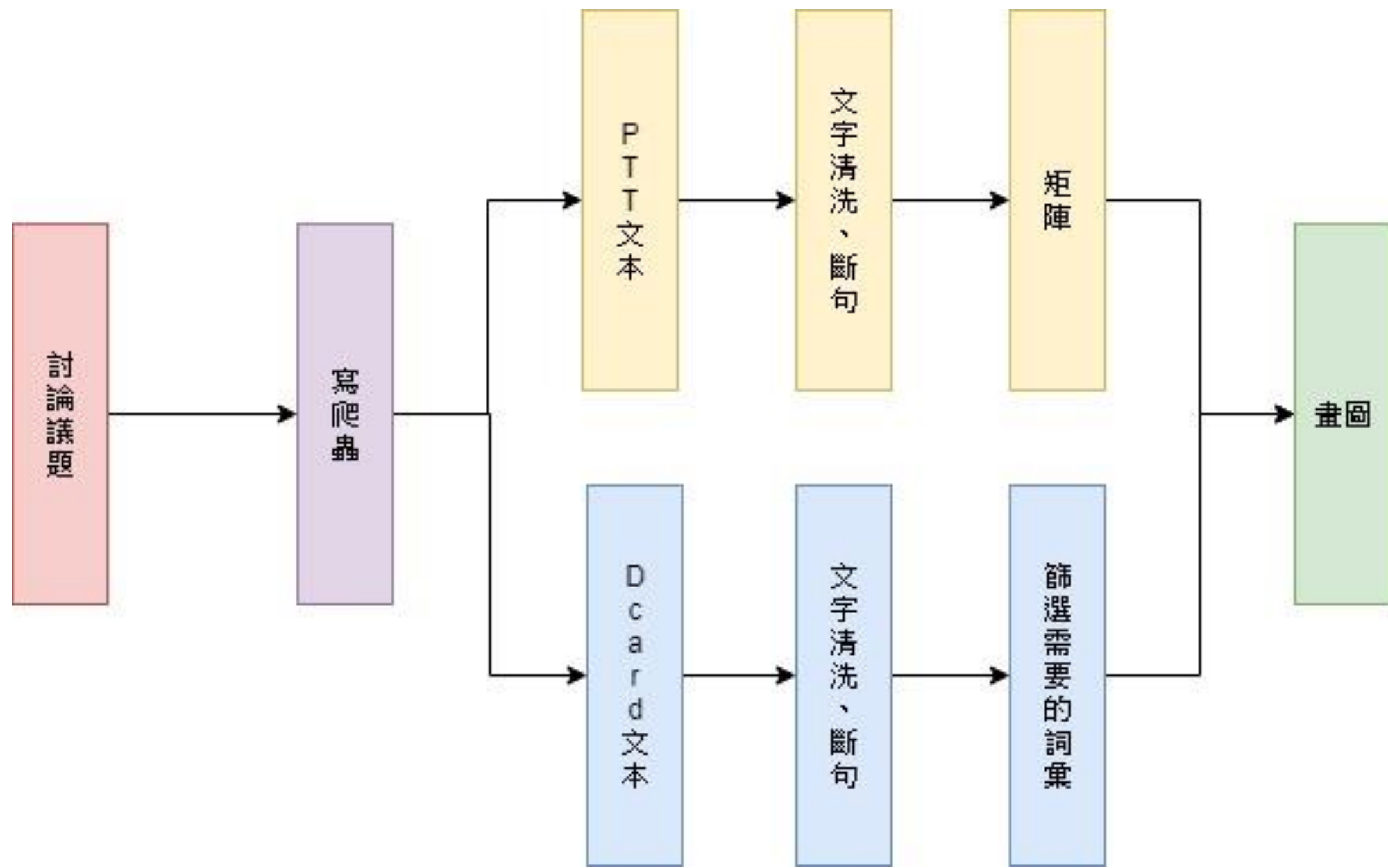


爬了90000筆資料
結果IP被Dcard Ban掉



一開始選太多關鍵字，
導致資料看起來雜亂。

分析流程



未來走向

- 刺激台灣人思考「一國兩制」是否可行
- 影響2020總統大選，台灣社會對中國疑慮加深
- 對立場強調台灣主體性的蔡英文總統形成政治助力