**社群媒體分析**

**第一次讀書會報告\_第十二組**

**一、分析主題**

民眾對星宇航空網路輿情探析

**二、組員**

N114320020\_曾子瑋

N114320008\_陳家卉

N114320017\_黃敏喩

N114320031\_洪宜綾

N114320002\_李湘怡

N114320027\_吳政翰

N114320030\_吳郁文

**三、分析工具**

* 中山大學工作流程平台
* 工作流程名稱：第12組\_第一次讀書會報告

**四、動機**



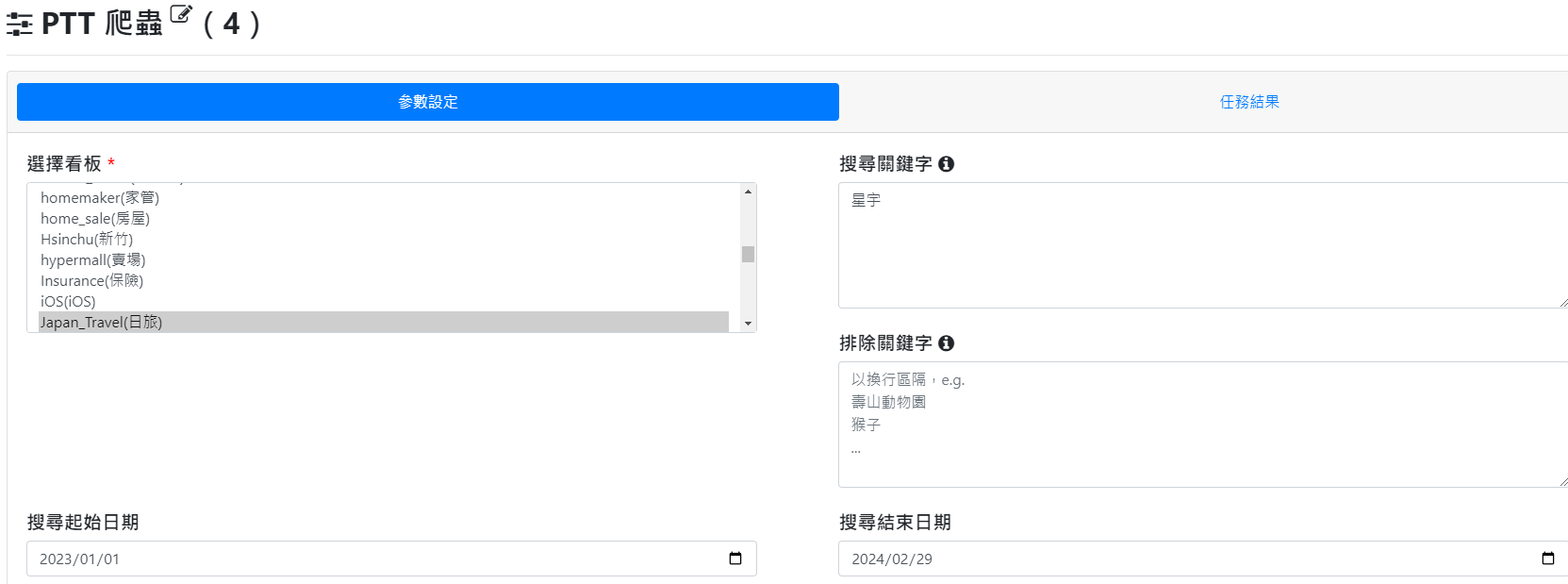
星宇航空成立於2020年，從尚未正式成立時董事長張國煒所屬長榮集團上演的家族遺產內鬥就吸引了大量的目光，而後張國煒更是上演了一齣王子復仇記的戲碼下，成立了星宇航空。

故此星宇航空便格外地受到矚目，而成立後不久更受到大環境武漢肺炎疫情影響，星宇航空透過從品牌定位形塑差異化優勢，如以目標服務客群市場區隔，鎖定未被滿足的高階精品航空服務需求客戶、高質感路線的質感與數位化視覺化宣傳方式，做出品牌形象，從價格兢爭激烈的航空市場中，以新進入市場競爭者竄起並擁有一定市占率，故引發本小組想了解星宇航空在社群網路上普遍情緒。

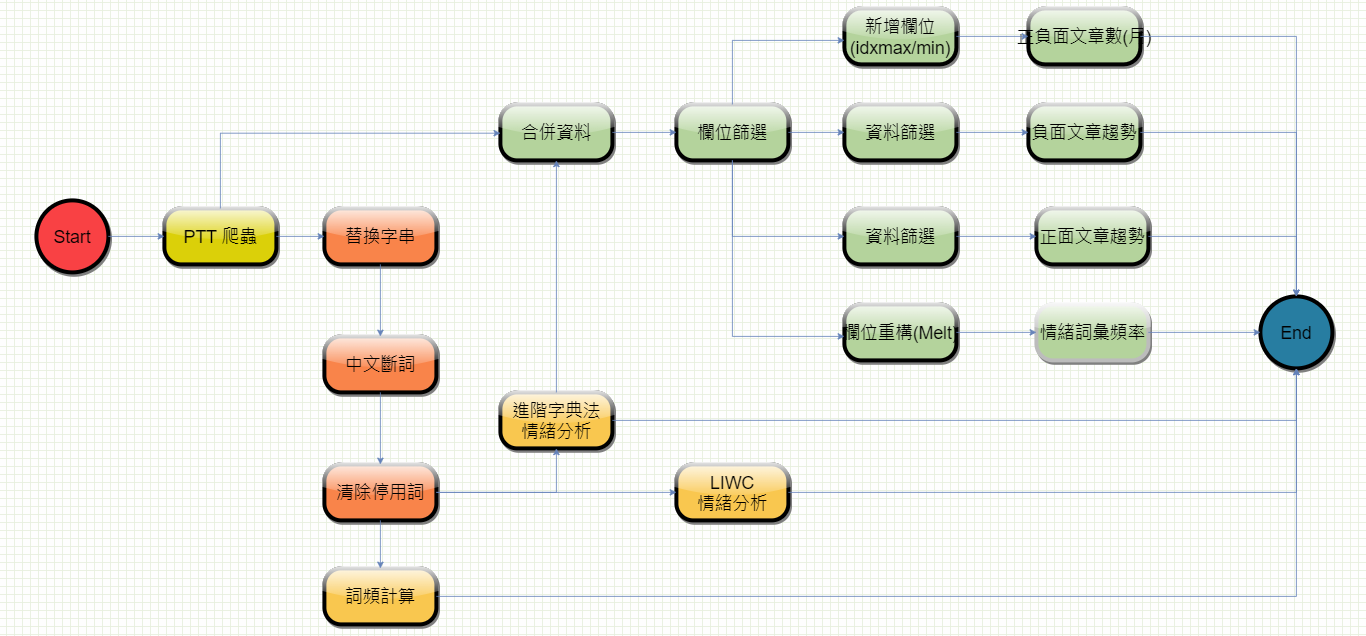
**五、資料範圍、來源、關鍵字**

* 資料來源：
  + PTT八卦版(Gossiping)
  + PTT航空板(Aviation)
  + PTT日旅(Japan\_Travel )
* 資料範圍：2023/01/01~2024/02/29
* 搜尋關鍵字：星宇

## **六、PTT爬蟲任務結果**



## **七、系統流程**



**八、分析過程**

**8-1 替換字串**

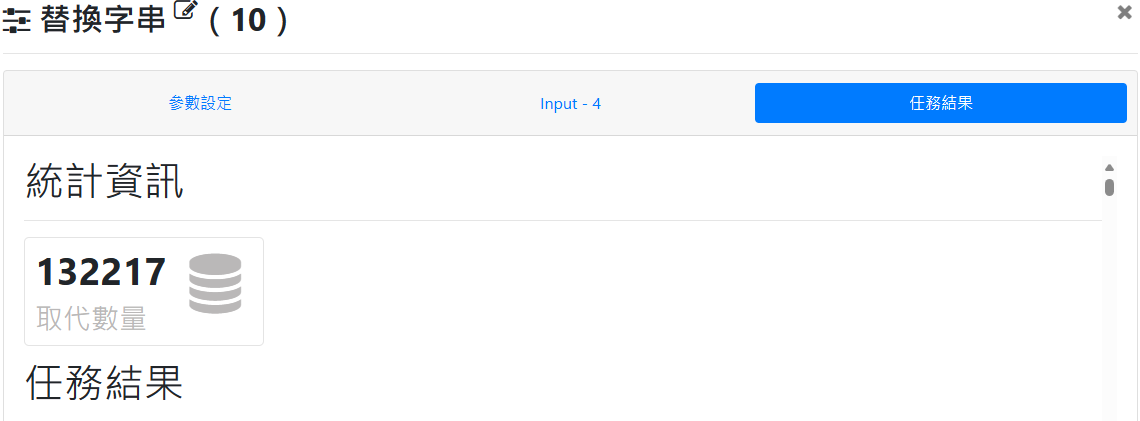
8-1-1 將文章內"\n\n">>換成''。''

8-1-2 將文章內"\n">>換成''，''

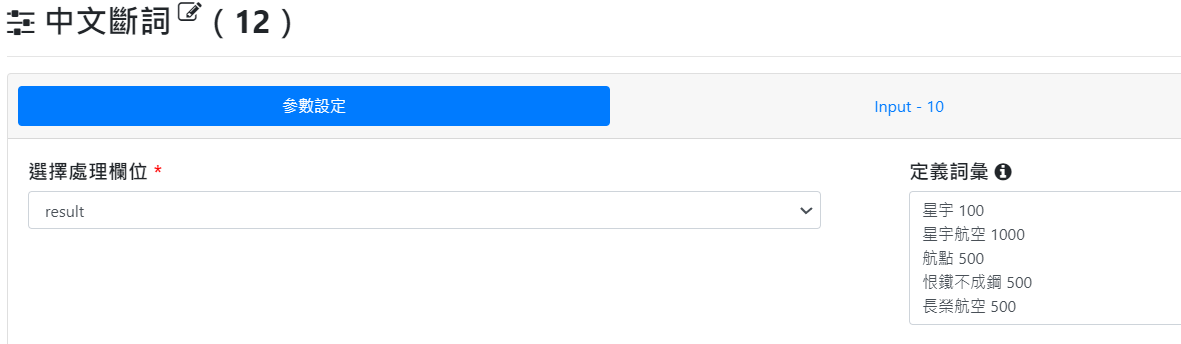
8-1-3 透過正規表達式將網址、網域、IP位址、路徑等資訊刪除((http|ftp|https)://)(([a-zA-Z0-9\.\_-]+\.[a-zA-Z]{2,6})|([0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}))(:[0-9]{1,4})\*(/[a-zA-Z0-9\&%\_\./-~-]\*)?>>  
8-1-4將文章內"Sent from JPTT on my \w+">> 更換成''空白''



8-1-4 任務結果: 總共替換了132217個字串



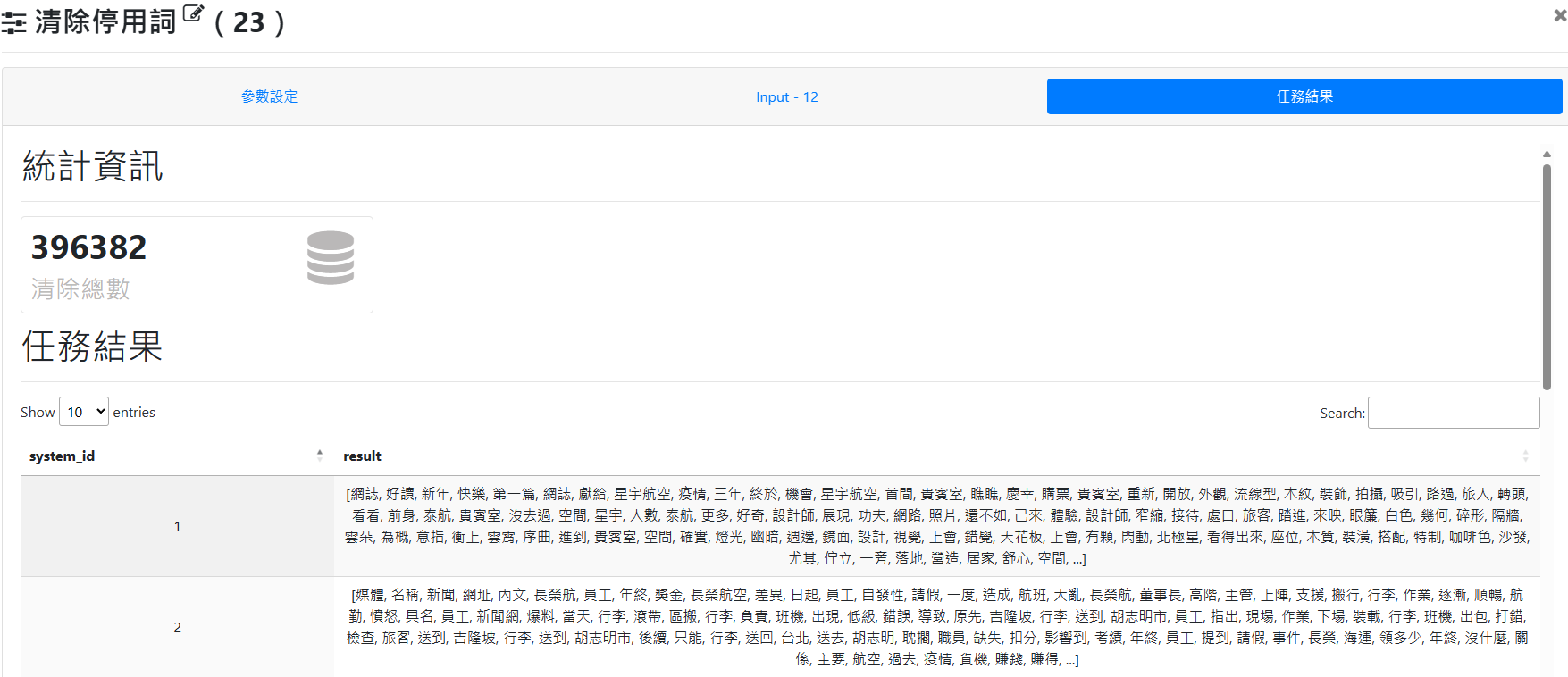
**8-2中文斷詞設定**

8-2-1參數設定:星宇少部分斷詞還是會出現“星””宇“；星宇航空會斷詞為"星宇""航空"；長榮航空會斷詞為"長榮""航空"；航點會斷詞為"航""點"；恨鐵不成鋼會斷詞為"恨""鐵""不成""鋼"，因此增加權重星宇100、星宇航空1000、長榮航空500、航點500、恨鐵不成鋼500

8-2-2清除停用詞參數設定：自定義停用字41個

“發現”“預計”“前往”“知道”“完成”“日期”“決定”“朋友”“情人”“確定”“需要”“特別”“今天”“看到”“八卦”“寶貝”“溫泉”“早上”“晚上”"之後”“現在”“一下”“提供”“小時”“分鐘”“投入”“表示”“直接”“使用”“最後”“目前”“人員”“準備”“時間””機上”“選擇”“搭乘” “改進”“最後”“表示”“有沒有”

8-2-3任務結果：清除396382個停用字

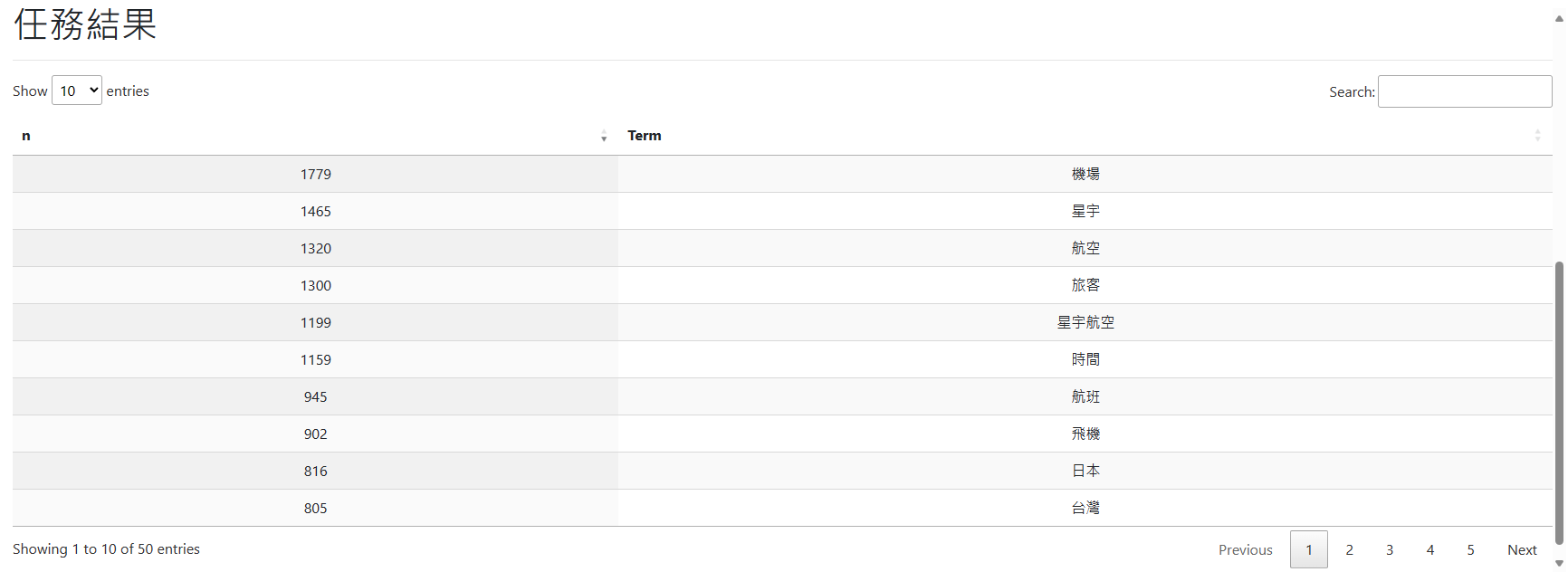
  
**8-3詞頻計算**

8-3-1 調整篩選詞彙至50個，並針對原先結果再增加「準備、機上、搭乘」停用字，確保關鍵字聚焦在星宇航空。



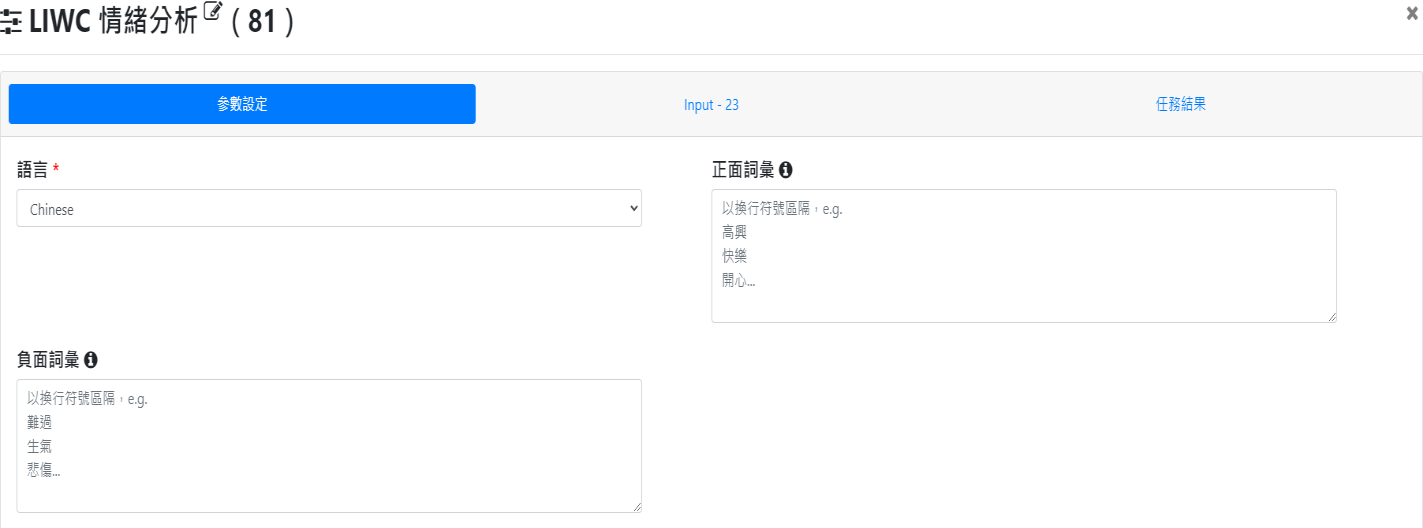


8-3-2任務結果

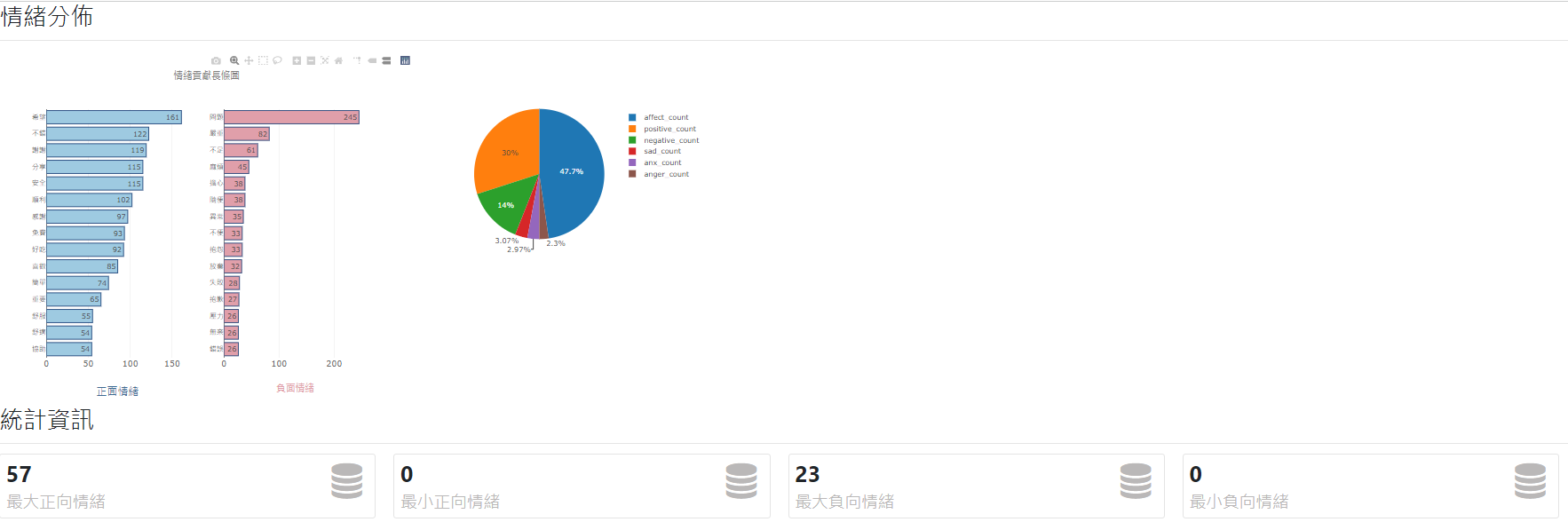


**8-4LIWC 情緒分析**

8-4-1參數設定： 設定語言為中文。

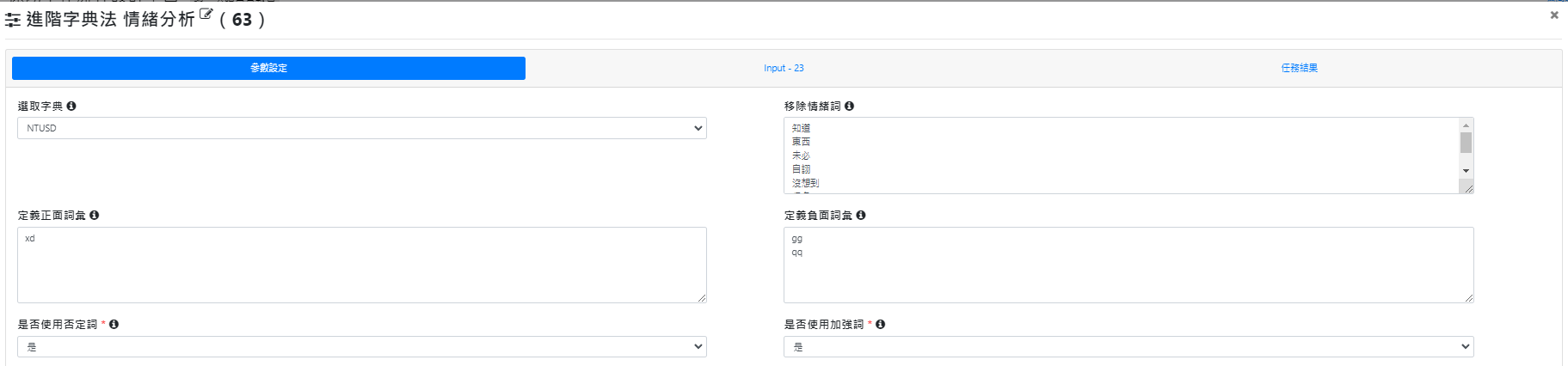


8-4-2任務結果：最大正向情緒數57，最大負向情緒數23。



**8-5進階字典法情緒分析**

8-5-1參數設定：選擇NTUSD作為情緒字典



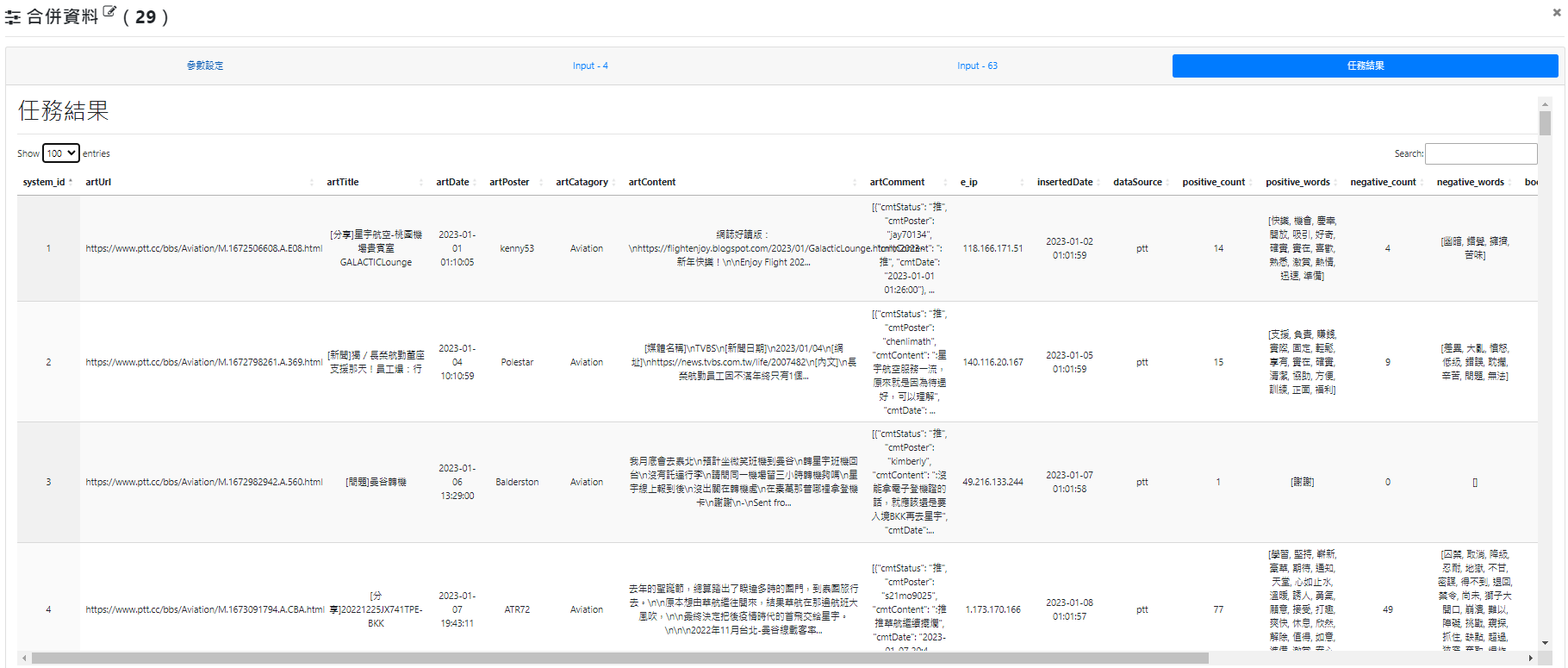
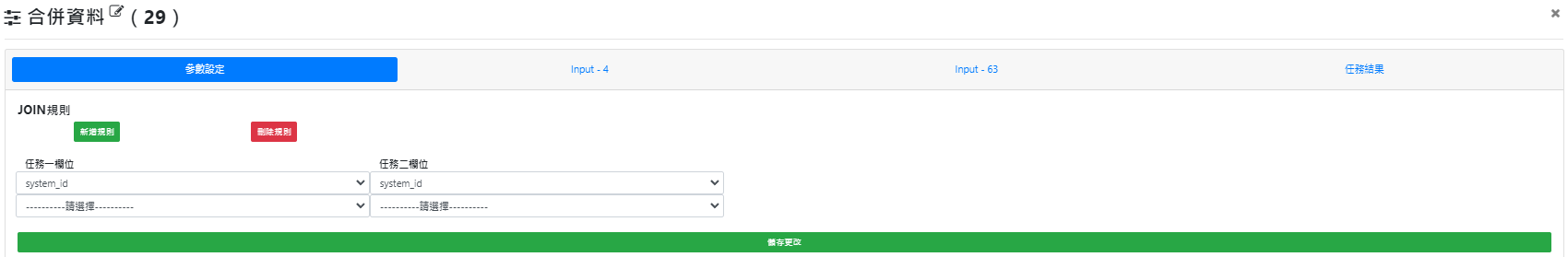
8-5-2移除知道、東西、未必、自詡 、沒想到、很多、希望、不到以上這些情緒詞，並且將xd 定義為正面詞彙， gg、qq 定義負面詞彙。

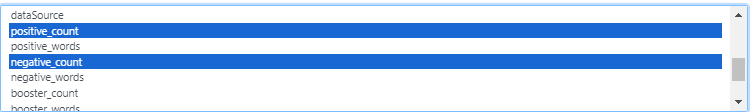
8-5-3任務結果



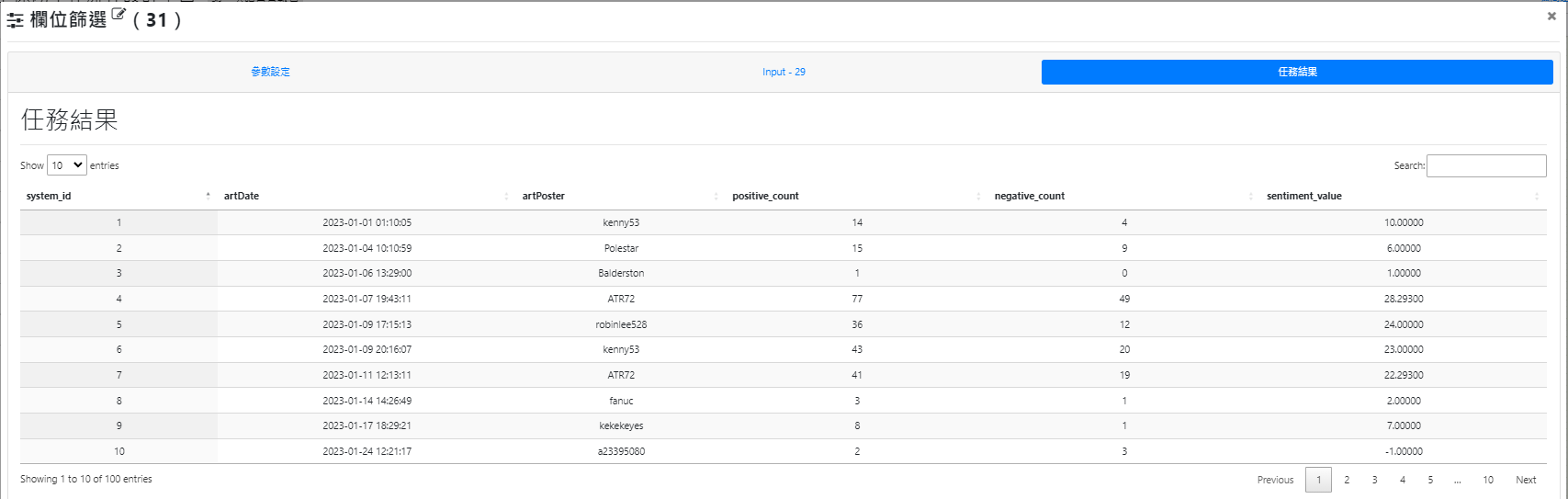
**8-6合併資料**

因要分別進行正負面情緒分析,利用已分析出來的進階字典法情緒分析與PTT爬蟲來源，以system\_id欄位做為key值，進行join。

**8-7欄位篩選**

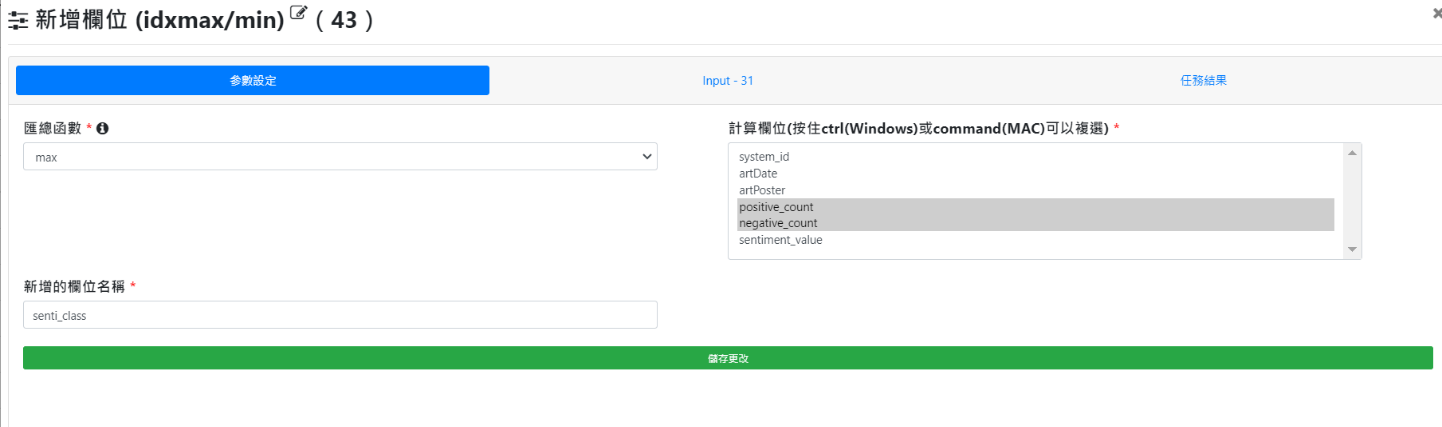
8-7-1參數設定：利用 system\_id、artDate、artPoster、positive\_count、negative\_count及sentiment\_Value六個欄位，進行時間與正負面情緒分析 

# 8-7-2任務結果

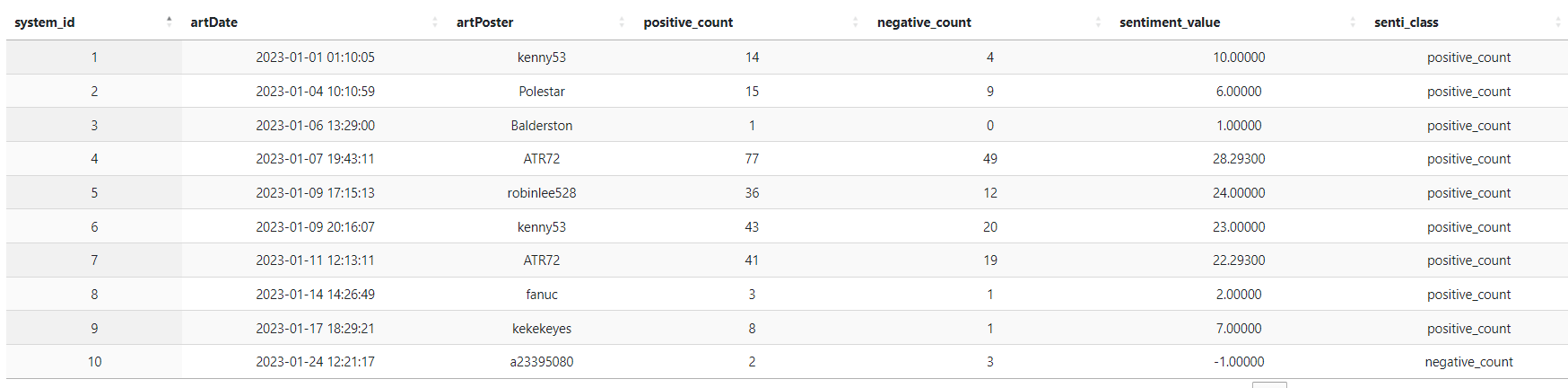


**8-8新增欄位**

8-8-1為了要判斷文章中是屬於正面文章(正面分數>負面分數)或負面文章 (正面分數<負面分數)，需要先新增一個判定欄位 senti\_class



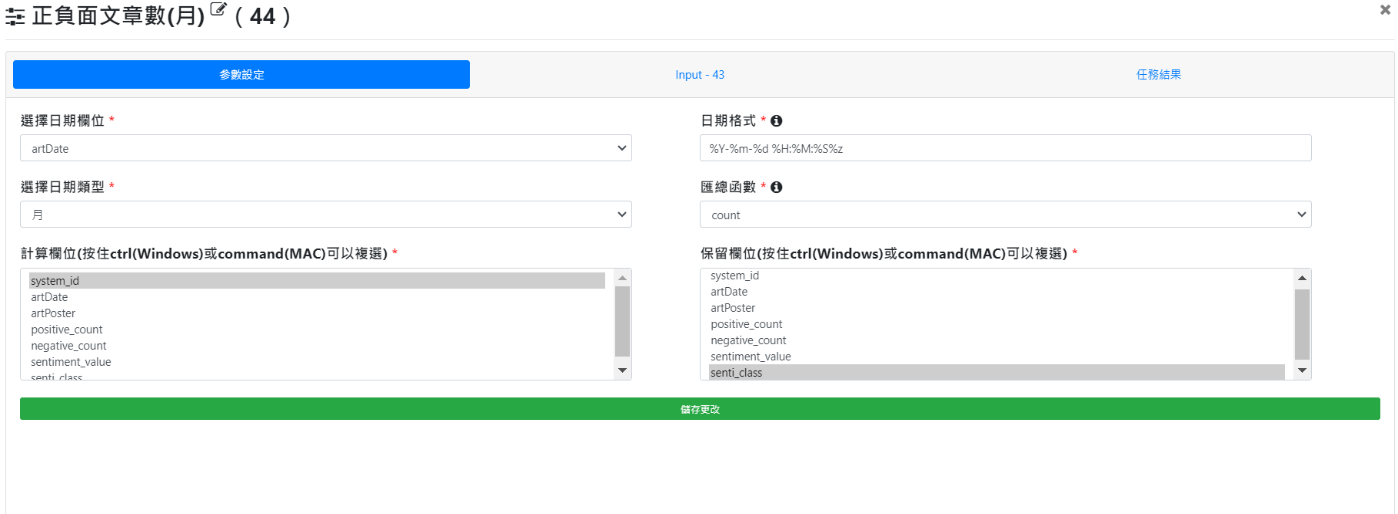
8-8-2任務結果：詳下圖最右方欄位。

**8-9分群彙總**

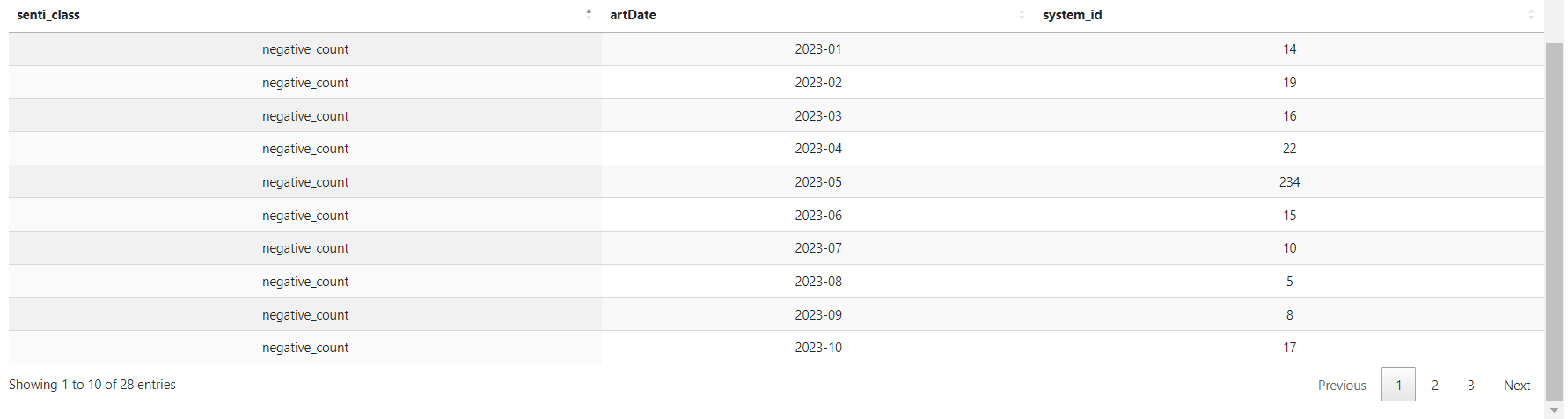
8-9-1正負面文章數

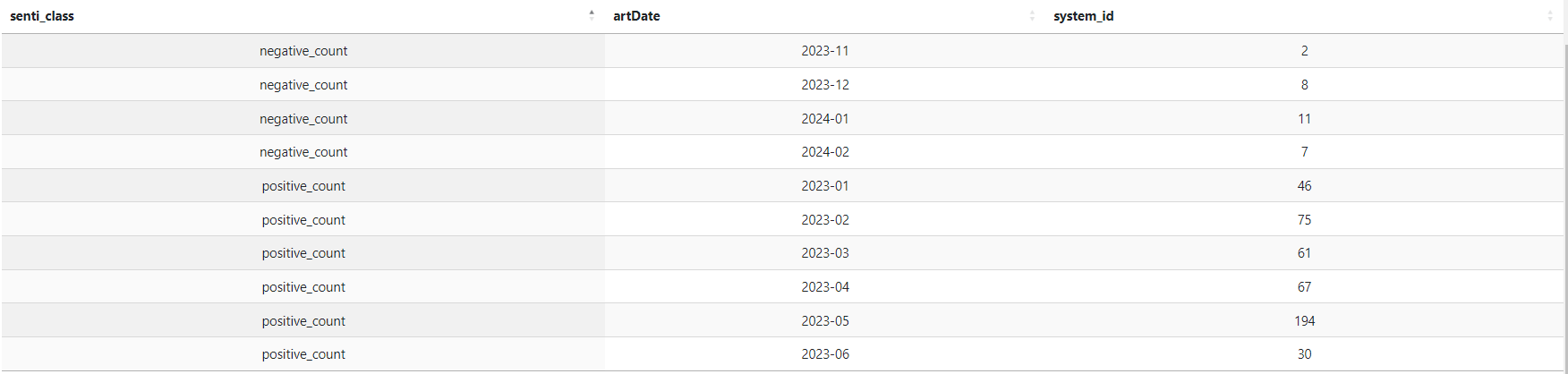
8-9-2使用分群彙總功能計算每月正負面文章數量。

8-9-3參數設定：選擇日期類型為"月"、彙總函數為"count”、計算欄位選system\_id、保留欄位選 senti\_class。



8-9-4任務結果：分每月正負面文章數量，呈現表格如下。

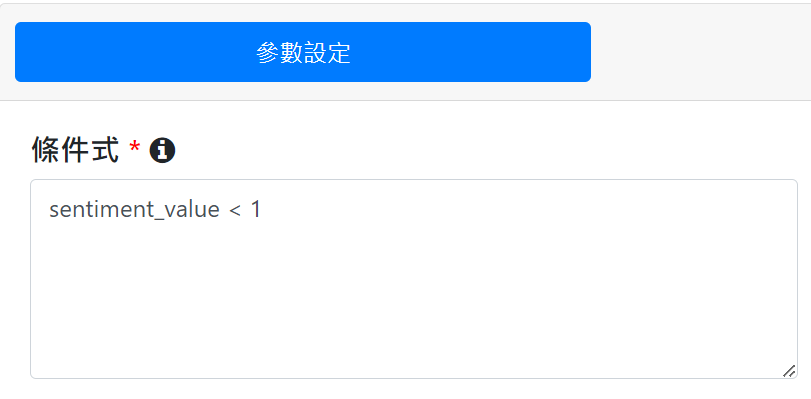




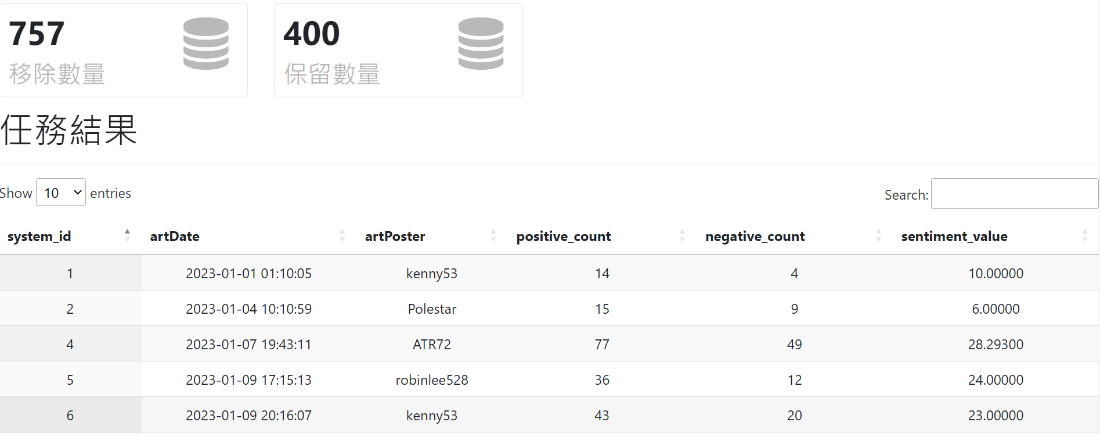
**8-10資料篩選**

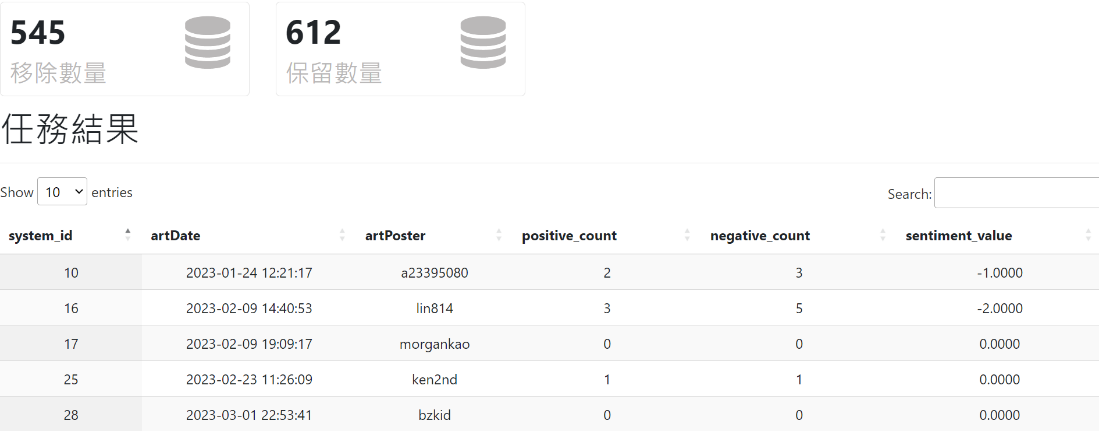
8-10-1參數設定：正面情緒條件式為” sentiment\_value > 1”; 負面情緒為” sentiment\_value < 1”。





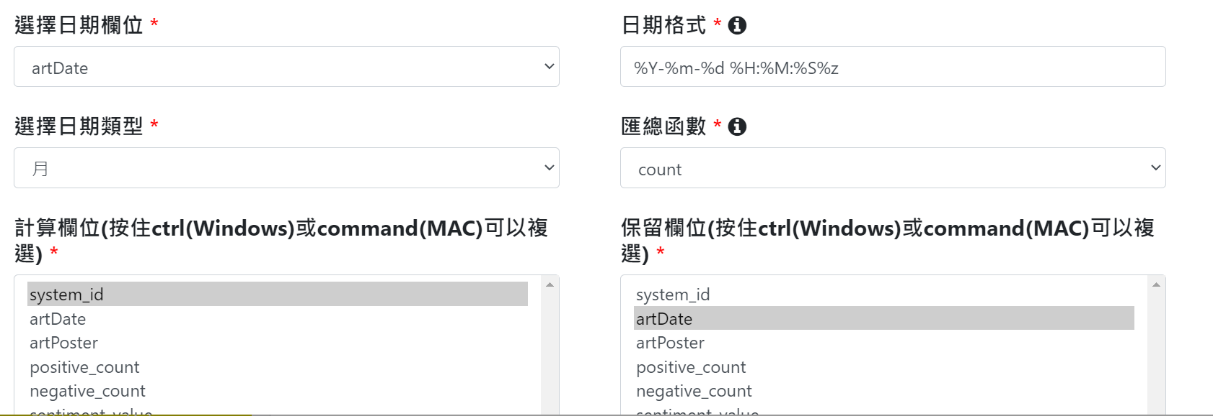
8-10-2任務結果:分別400筆正向情緒及612筆負向情緒資料。



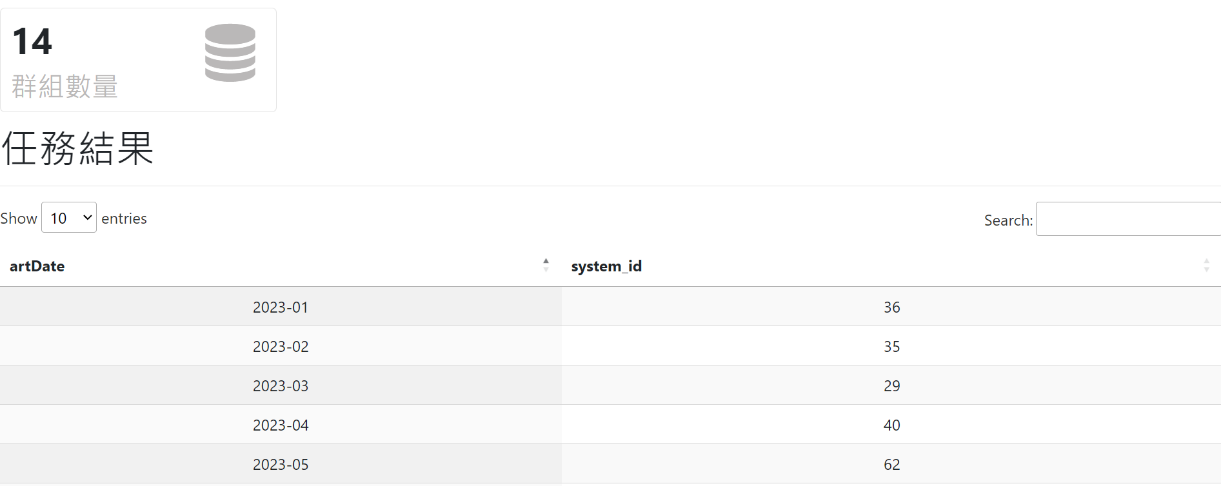


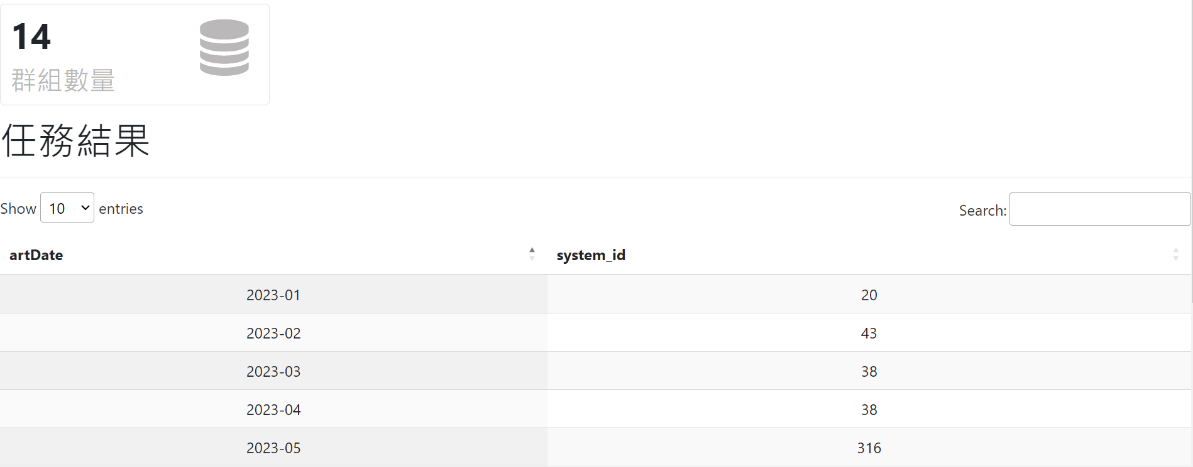
**8-11正(負)面情緒分析**

8-11-1參數設定：合併匯總group by 日期artDate,，並選擇以月份為單位，匯總函數用count去計算篇數(system\_id)，結果可看到各月份有正(負)面情緒的文章篇數。



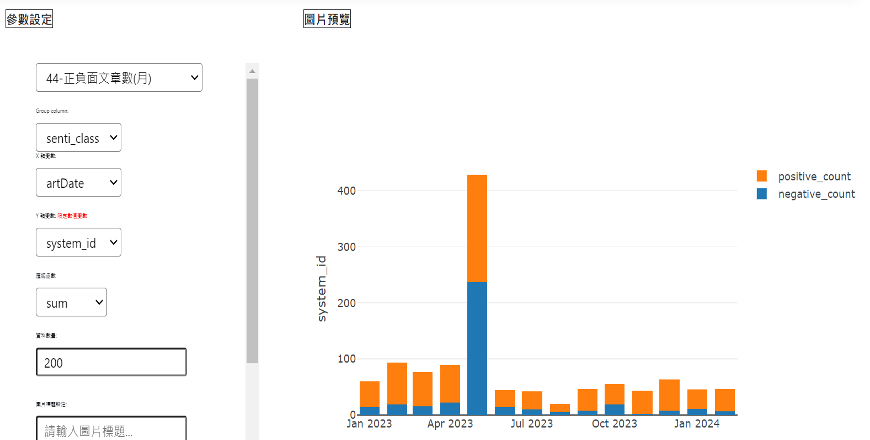
8-11-2任務結果：





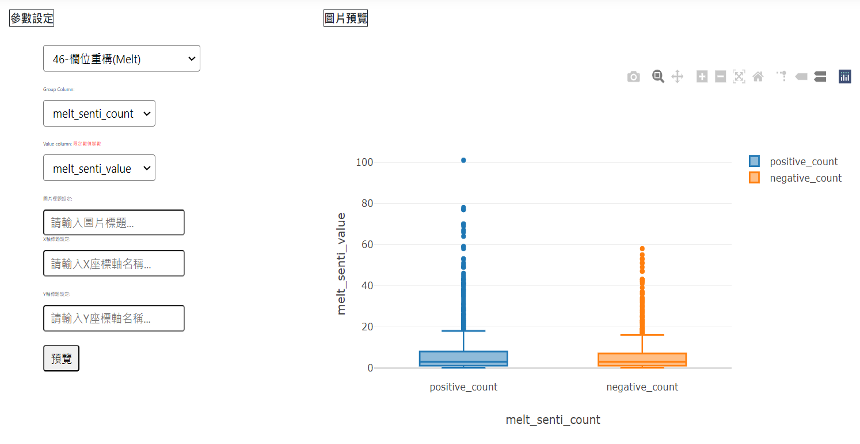
## **九、儀表板-直條圖**

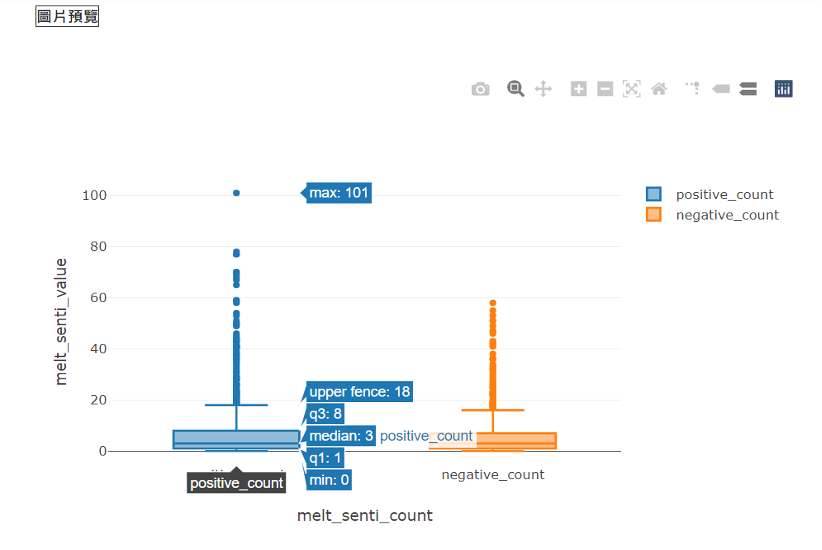
* 利用直條圖可以看到不同時間，網民對星宇航空的情緒反應。
* 結論：在2023年五月份星宇發生「在東京成田機場因為遇上天候因素，導致延誤滯留7小時，加上遇到機場宵禁，最後302名旅客被迫在機場過夜」的負面新聞以後，五月份網民負面情緒分數明顯比正面情緒高，其他月份則均為正面>負面，且整體討論熱度也明顯高於其他月份高。

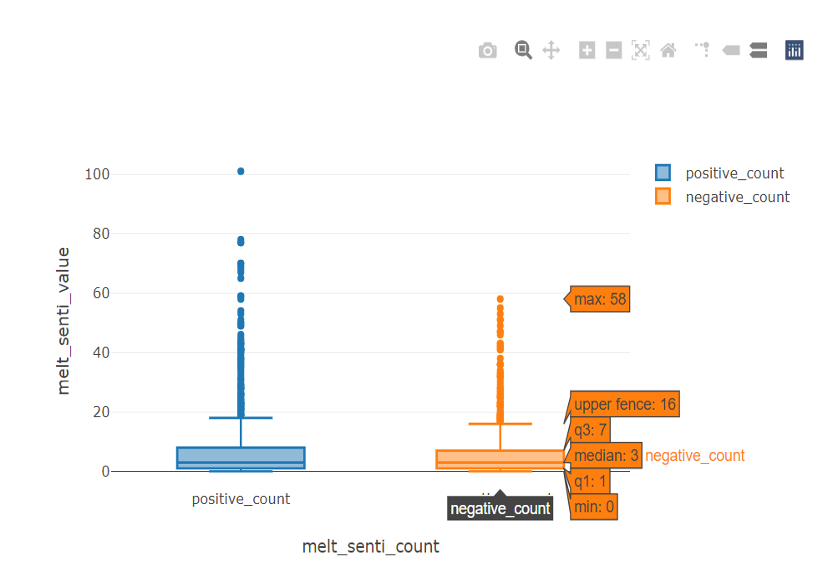


## **十、儀表板-箱型圖**

* 利用箱型圖來看正面情緒與負面情緒的資料分布。
* 正面與負面情緒的盒狀都很短，代表兩者的四分位數範圍較小且數據集中， 顯示平常對於星宇航空的情緒值變動不大，但離群值卻很多，正面離群值最大101，負面離群值最大58，推測可能在研究期間內若有關於星宇的新聞出現時，會出現大量的討論所導致。此外，針對情緒分數較高的po文者，本組有特別至ptt航空板搜尋該作者，發現是航空迷所發的長篇搭乘體驗文，因此情緒分數會特別高。  
  [[分享] 20221225 JX741 TPE-BKK](https://moptt.tw/p/Aviation.M.1673091794.A.CBA#google_vignette) [[分享] 20230305 JX745 TPE-BKK](https://moptt.tw/p/Aviation.M.1679236564.A.7BA#google_vignette)

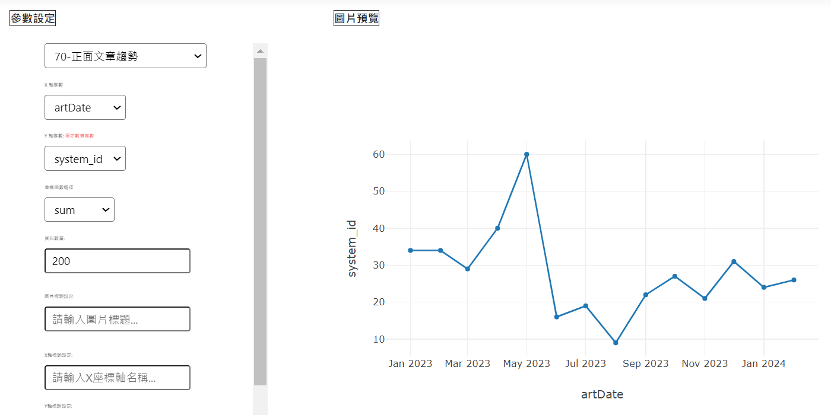
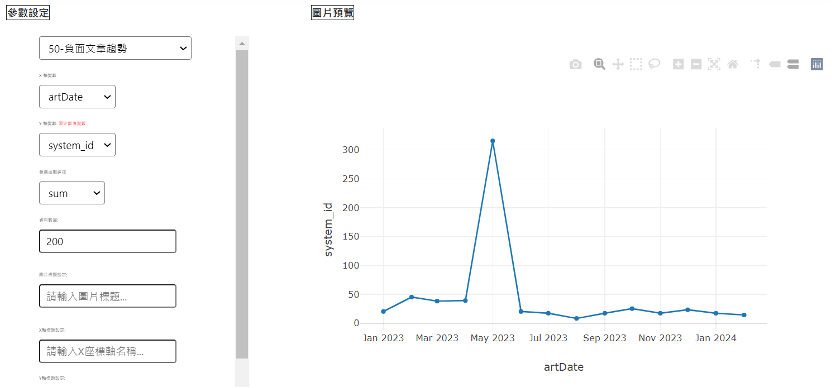






**十一、儀表板-折線圖**

* 利用折線圖來看正負向情緒的波動。
* 很明顯地，2023年五月份時正負面的情緒波動都呈現激烈的波動，討論熱度很高，其他月份正負向情緒文約當在40則。



**十二、結論**

* 利用PTT八卦板、航空版與日旅版，觀察網友對於星宇航空的看法。
* 除了發生東京延誤事件以外的其他月份，以直條圖來看，網路對於星宇航空的評價大部分是正面大於負面。但於2023/5月份發生「在東京成田機場因為遇上天候因素，導致延誤滯留7小時，加上遇到機場宵禁，最後302名旅客被迫在機場過夜」的負面新聞以後，當月份整體討論熱度升高，而且負面情緒很明顯高於正面情緒，顯示網民對於星宇航空的處理態度並不滿意。
* 但過了2023/5月份的新聞熱點以後，又回歸到正面高於負面情緒的狀態，顯示雖然星宇航空在該事件的危機處理並未讓大眾滿意，但平常時的服務表現應該還是在不錯的水準，該事件對於星宇的長久營運來說並無重大影響。