

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_黃敏喻

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，從價格競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

PTT 爬蟲 (4)

參數設定

任務結果

選擇看板 *

homemaker(家管)

home_sale(房產)

Hsinchu(新竹)

hypermall(賣場)

Insurance(保險)

iOS(iOS)

Japan_Travel(日旅)

搜尋關鍵字 ❶

星宇

排除關鍵字 ❶

以換行區隔，e.g.

泰山動物園

獅子

...

搜尋起始日期

2023/01/01

搜尋結束日期

2024/02/29

PTT 爬蟲 (4)

參數設定

統計資訊

10

欄位數

1157

資料筆數

☰ 替換字串  (10)

參數設定

Input - 4

任務結果

選擇處理欄位 *

artContent

替換字串設定 ⓘ

```
\n\n>> °  
\n>> ,  
((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]{1,4}))*\/[a-zA-Z0-9\&%_\~\-\*]*)?>>
```

選擇替換規則檔案 ⓘ

-----請選擇-----


儲存更改

8-1-4 任務結果: 總共替換了 132217 個字串

≡ 替換字串 (10)

參數設定 Input - 4 任務結果

統計資訊

132217 

取代數量

任務結果

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

≡ 中文斷詞 (12)

參數設定

Input - 10

選擇處理欄位 *

result

定義詞彙 ⓘ

星字 100
星宇航空 1000
航點 500
恨鐵不成鋼 500
長榮航空 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

清除停用詞

(23)

參數設定

Input - 12

任務結果

統計資訊

396382

清除總數

任務結果

Show 10 entries

Search:

system_id	result
1	[副話, 好聽, 新年, 快樂, 第一輯, 副話, 獻給, 皇宇航空, 疫情, 三年, 終於, 機會, 皇宇航空, 普聞, 貴賓室, 曉曉, 雲朵, 購買, 貴賓室, 重新, 開站, 外觀, 流線型, 木紋, 裝飾, 拍攝, 吸引, 踏進, 旅人, 精神, 看看, 前身, 泰航, 貴賓室, 沒去過, 空間, 量字, 人數, 泰航, 更多, 好奇, 設計師, 展現, 功夫, 網絡, 圖片, 還不如, 已來, 體驗, 設計師, 至極, 接待, 處口, 旅客, 踴躍, 來說, 瞭解, 白色, 餐椅, 碎形, 隔牆, 雲朵, 為嘅, 意指, 畫上, 雲霄, 序曲, 進到, 貴賓室, 空間, 確實, 燈光, 幽暗, 環境, 橫面, 設計, 視覺, 上會, 錯覺, 天花板, 上會, 有顆, 閃動, 北極星, 看得出來, 座位, 木質, 裝潢, 搭配, 特別, 咖啡色, 沙發, 尤其, 佇立, 一旁, 落地, 簾造, 居家, 舒心, 空間...]
2	[媒體, 名稱, 新聞, 網址, 內文, 長榮航, 員工, 年終, 獎金, 長榮航空, 差異, 日誌, 員工, 自發性, 請假, 一度, 造成, 航班, 大亂, 長榮航, 董事長, 高階, 主管, 上陣, 支援, 飛行, 行李, 作業, 逐漸, 順暢, 航勤, 憤怒, 異名, 員工, 新聞網, 爆料, 當天, 行李, 滾落, 區間, 行李, 負責, 班機, 出現, 低級, 錯誤, 導致, 誤失, 吉隆坡, 行李, 送到, 胡志明市, 員工, 指出, 現場, 作業, 下場, 裝載, 行李, 班機, 出包, 打錯, 檢查, 旅客, 送到, 吉隆坡, 行李, 送到, 胡志明市, 後機, 只能, 行李, 送回, 台北, 送去, 胡志明, 耽擱, 職員, 缺失, 扣分, 影響到, 老練, 年終, 員工, 提到, 請假, 事件, 長榮, 海運, 候多少, 年終, 沒什麼, 關係, 主要, 航空, 過去, 疫情, 災禍, 賺錢, 釋得...]

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。

詞頻計算 (26)

參數設定

Input - 23

任務結果

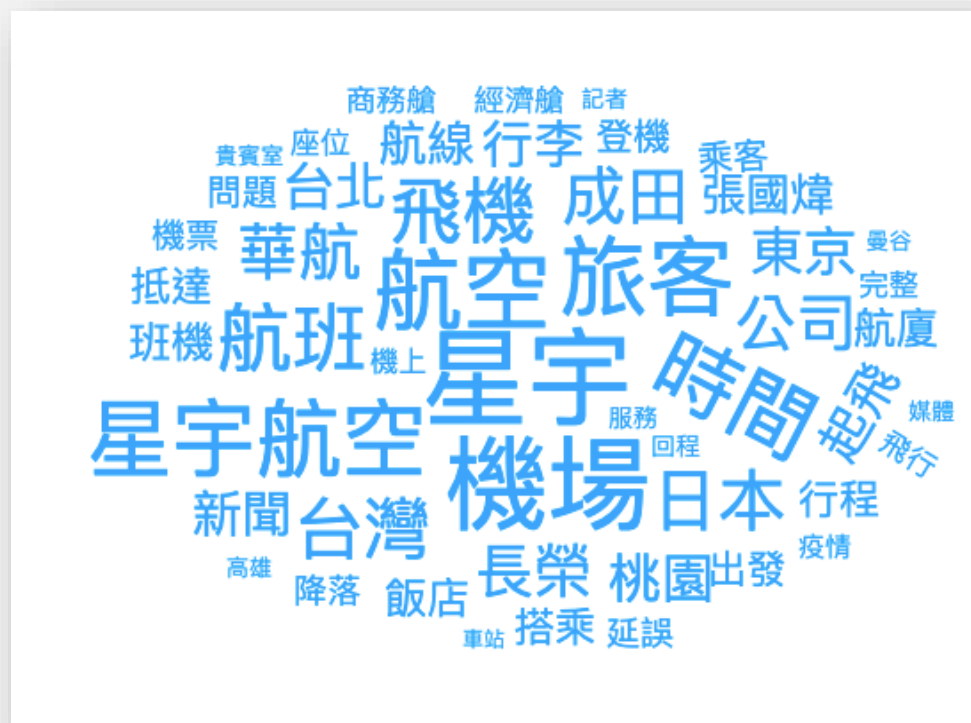
自定義停止詞

準備
機上
搭乘

最多篩選詞彙數量 1

50

儲存更改



8-3-2 任務結果

任務結果

Show entries Search:

n	Term
1779	機場
1465	星宇
1320	航空
1300	旅客
1199	星宇航空
1159	時間
945	航班
902	飛機
816	日本
805	台灣

Showing 1 to 10 of 50 entries Previous **1** 2 3 4 5 Next

8-4LIWC 情緒分析

8-4-1 參數設定：設定語言為中文。

LIWC 情緒分析 (81)

參數設定 Input - 23 任務結果

語言 *

Chinese

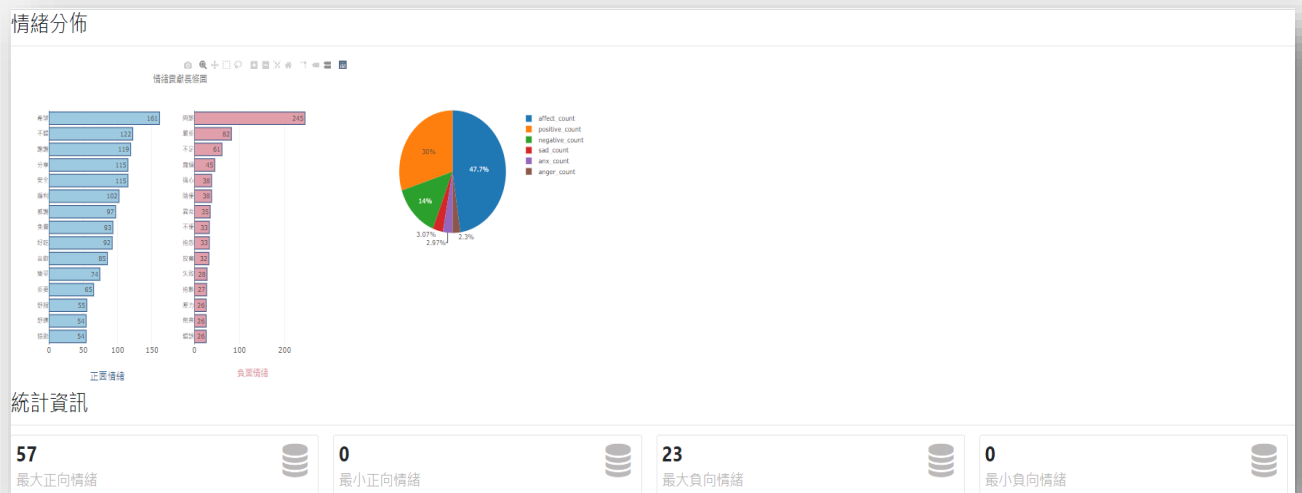
正面詞彙 ⓘ

以換行符號區隔，e.g.
高興
快樂
開心...

負面詞彙 ⓘ

以換行符號區隔，e.g.
難過
生氣
悲傷...

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

進階字典法 情緒分析 (63)

參數設定

Input - 23

任務結果

選取字典 0

NTUSD

定義正面詞彙 0

xd

是否使用否定詞 0

是

移除值選詞 0

知道
東西
未必
自詡
沒想到

定義負面詞彙 0

gg
qq

是否使用加強詞 0

是

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。

合併資料 (29)

參數設定 Input-4 Input-63 任務結果

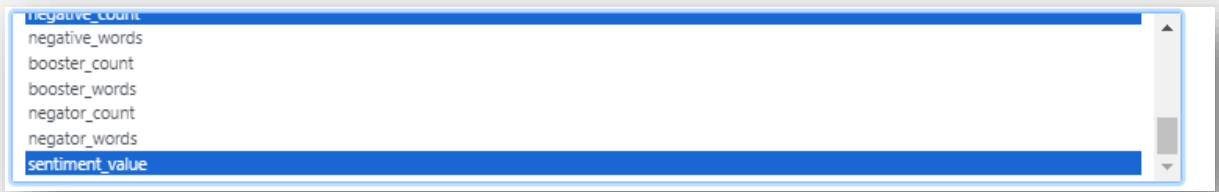
JOIN規則

新增規則 刪除規則

任務一欄位 任務二欄位

system_id	system_id
.....請選擇.....請選擇.....

儲存更改



8-7-2 任務結果

欄位篩選 (31)

參數設定 Input - 29 任務結果

任務結果

Show 10 entries

system_id	artDate	artPoster	positive_count	negative_count	sentiment_value
1	2023-01-01 01:10:05	kenny53	14	4	10.00000
2	2023-01-04 10:10:59	Polestar	15	9	6.00000
3	2023-01-06 13:29:00	Balderston	1	0	1.00000
4	2023-01-07 19:43:11	ATR72	77	49	28.29300
5	2023-01-09 17:15:13	robinlee528	36	12	24.00000
6	2023-01-09 20:16:07	kenny53	43	20	23.00000
7	2023-01-11 12:13:11	ATR72	41	19	22.29300
8	2023-01-14 14:26:49	fanuc	3	1	2.00000
9	2023-01-17 18:29:21	kekekeyes	8	1	7.00000
10	2023-01-24 12:21:17	a23395080	2	3	-1.00000

Showing 1 to 10 of 100 entries

Previous 1 2 3 4 5 ... 10 Next

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

新增欄位 (idxmax/min) (43)

參數設定 Input - 31 任務結果

匯總函數 * 1

max

計算欄位(按住ctrl(Windows)或command(MAC)可以複選) *

system_id
artDate
artPoster
positive_count
negative_count
sentiment_value

新增的欄位名稱 *

senti_class

儲存更改

8-8-2 任務結果：詳下圖最右方欄位。

system_id	artDate	artPoster	positive_count	negative_count	sentiment_value	senti_class
1	2023-01-01 01:10:05	kenny53	14	4	10.00000	positive_count
2	2023-01-04 10:10:59	Polestar	15	9	6.00000	positive_count
3	2023-01-06 13:29:00	Balderston	1	0	1.00000	positive_count
4	2023-01-07 19:43:11	ATR72	77	49	28.29300	positive_count
5	2023-01-09 17:15:13	robinlee528	36	12	24.00000	positive_count
6	2023-01-09 20:16:07	kenny53	43	20	23.00000	positive_count
7	2023-01-11 12:13:11	ATR72	41	19	22.29300	positive_count
8	2023-01-14 14:26:49	fanuc	3	1	2.00000	positive_count
9	2023-01-17 18:29:21	kekekeyes	8	1	7.00000	positive_count
10	2023-01-24 12:21:17	az3395080	2	3	-1.00000	negative_count

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。

參數設定

Input - 43

任務結果

選擇日期欄位 *

artDate

選擇日期類型 *

月

計算欄位(按住ctrl(Windows)或command(MAC)可以複選) *

system_id

artDate

artPoster

positive_count

negative_count

sentiment_value

senti_class

日期格式 *

%Y-%m-%d %H:%M:%S%z

匯總函數 *

count

保留欄位(按住ctrl(Windows)或command(MAC)可以複選) *

system_id

artDate

artPoster

positive_count

negative_count

sentiment_value

senti_class

儲存更改

8-9-4 任務結果：分每月正負面文章數量，呈現表格如下。

senti_class	artDate	system_id
negative_count	2023-01	14
negative_count	2023-02	19
negative_count	2023-03	16
negative_count	2023-04	22
negative_count	2023-05	234
negative_count	2023-06	15
negative_count	2023-07	10
negative_count	2023-08	5
negative_count	2023-09	8
negative_count	2023-10	17

Showing 1 to 10 of 28 entries

Previous 1 2 3 Next

senti_class	artDate	system_id
negative_count	2023-11	2
negative_count	2023-12	8
negative_count	2024-01	11
negative_count	2024-02	7
positive_count	2023-01	46
positive_count	2023-02	75
positive_count	2023-03	61
positive_count	2023-04	67
positive_count	2023-05	194
positive_count	2023-06	30

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”。

參數設定

條件式 * ⓘ

sentiment_value > 1

參數設定

條件式 *

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。

757

移除數量

400

保留數量

任務結果

Show

10

 entries

Search:

system_id	artDate	artPoster	positive_count	negative_count	sentiment_value
1	2023-01-01 01:10:05	kenny53	14	4	10.00000
2	2023-01-04 10:10:59	Polestar	15	9	6.00000
4	2023-01-07 19:43:11	ATR72	77	49	28.29300
5	2023-01-09 17:15:13	robinlee528	36	12	24.00000
6	2023-01-09 20:16:07	kenny53	43	20	23.00000

545

移除數量

612

保留數量

任務結果

Show

10

 entries

Search:

system_id	artDate	artPoster	positive_count	negative_count	sentiment_value
10	2023-01-24 12:21:17	a23395080	2	3	-1.0000
16	2023-02-09 14:40:53	lin814	3	5	-2.0000
17	2023-02-09 19:09:17	morgankao	0	0	0.0000
25	2023-02-23 11:26:09	ken2nd	1	1	0.0000
28	2023-03-01 22:53:41	bzkid	0	0	0.0000

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。

選擇日期欄位 *

artDate

選擇日期類型 *

月

日期格式 * ⓘ

%Y-%m-%d %H:%M:%S%z

匯總函數 * ⓘ

count

計算欄位(按住ctrl(Windows)或command(MAC)可以複選) *

system_id
artDate
artPoster
positive_count
negative_count
sentiment_value

保留欄位(按住ctrl(Windows)或command(MAC)可以複選) *

system_id
artDate
artPoster
positive_count
negative_count
sentiment_value

8-11-2 任務結果：

14 群組數量

任務結果

Show 10 entries

Search:

artDate	system_id
2023-01	36
2023-02	35
2023-03	29
2023-04	40
2023-05	62

14

群組數量

任務結果

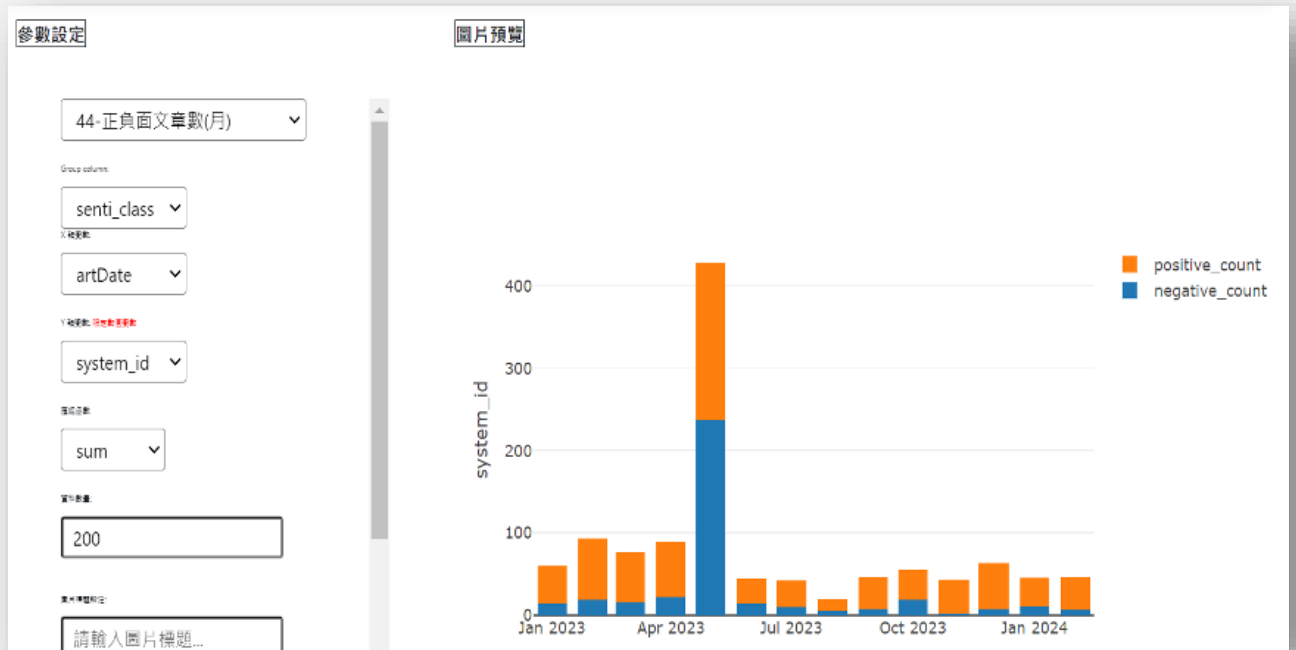
Show 10 entries

Search:

artDate	system_id
2023-01	20
2023-02	43
2023-03	38
2023-04	38
2023-05	316

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。

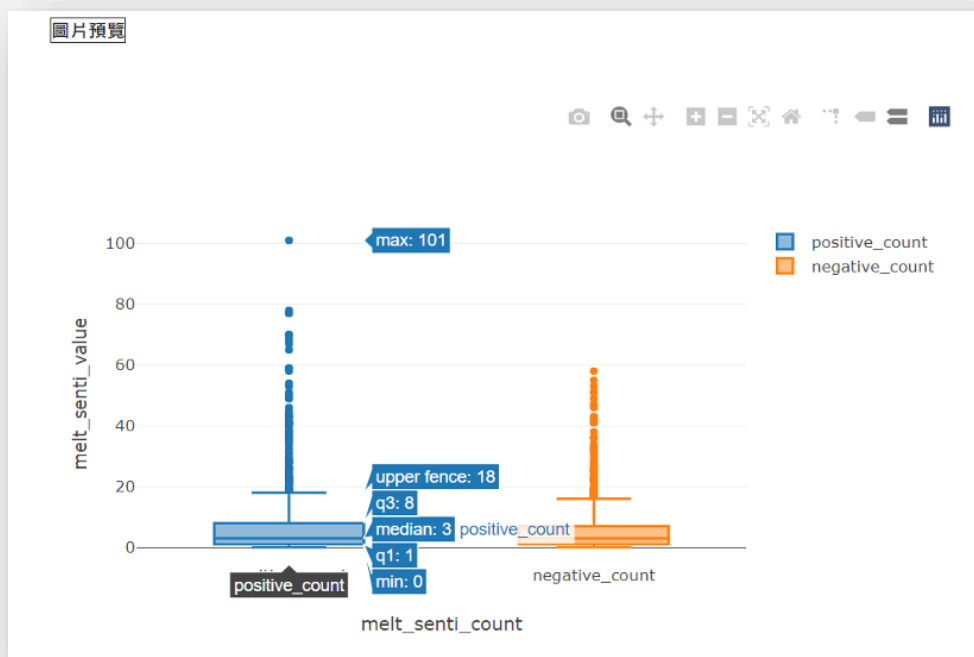
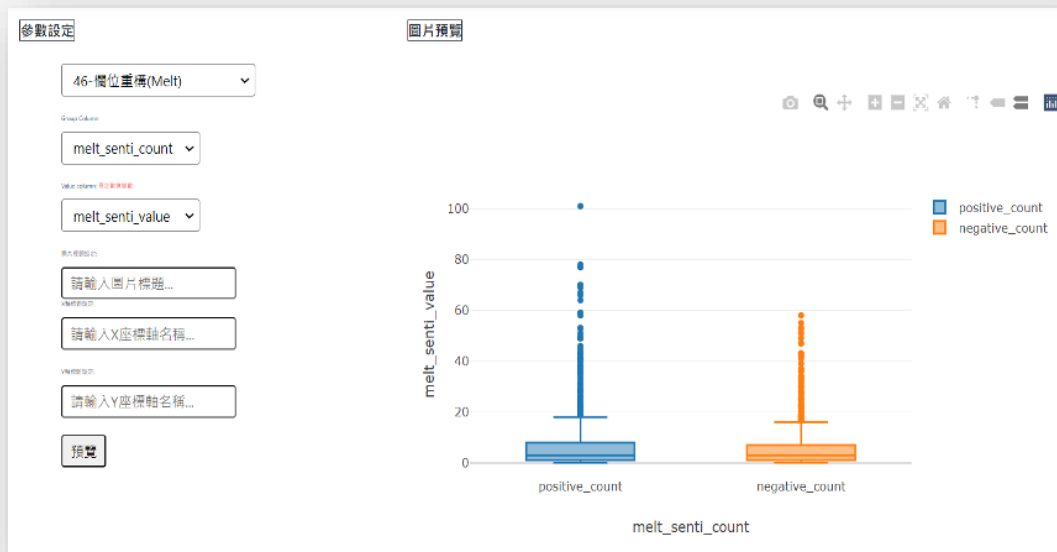


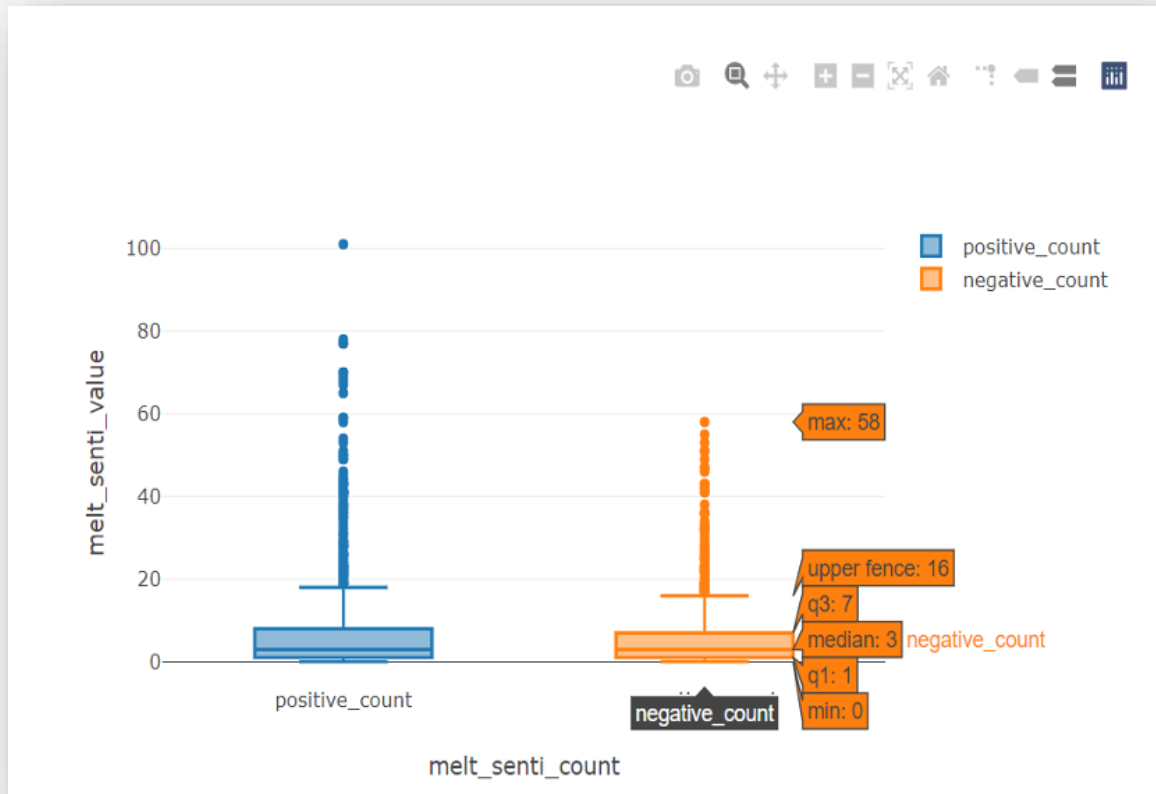
十、儀表板-箱型圖

- 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](#)

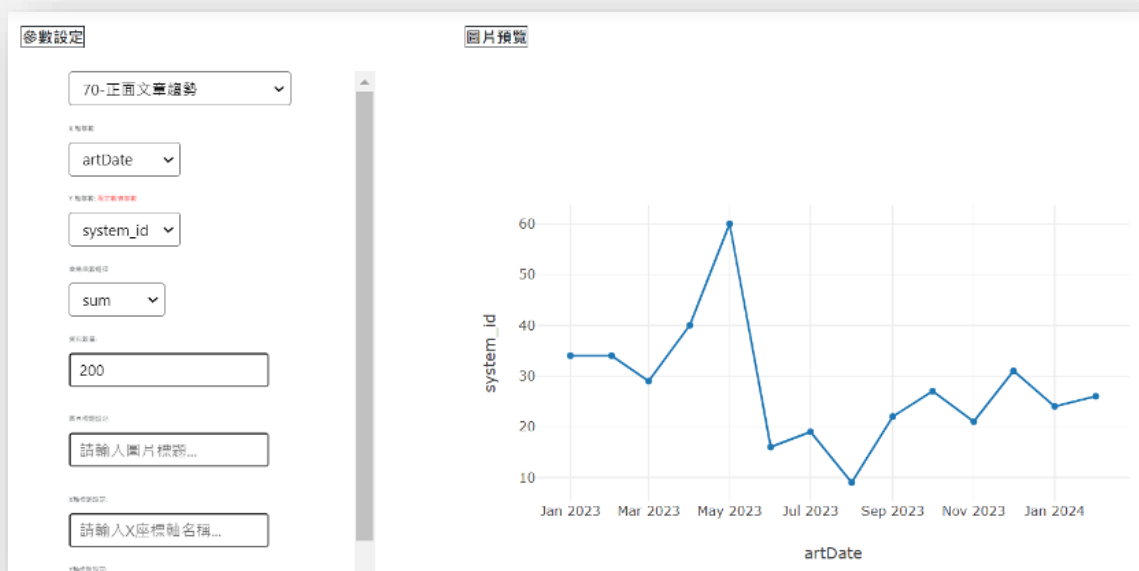
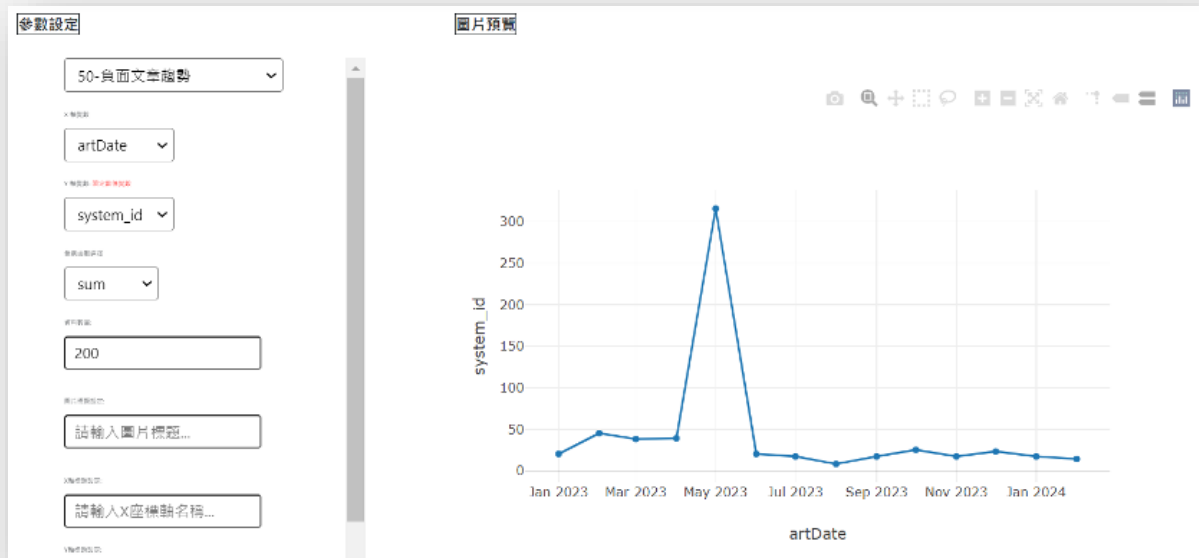
[\[分享\] 20230305 JX745 TPE-BKK](#)





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。