

第一次讀書會報告_第十三組

組員:

N124020001 林坤郁 N124020004 陳忠泰 N124020005 陳莘惠
N124020006 張祐倫 N124020012 郭展州

一、資料介紹

資料來源: Tarflow 平台/PTT 股票版(stock)文章

關鍵字: 台積電、美積電、亞利桑那

資料區間: 2024/9/1~2025/3/12，期間為時任美國總統參選人 Trump 於 2024 年 10/12 首次公開談話提到台灣偷走美國晶片產業至 2025 3/12 台積電董事長魏哲家於美國白宮正式宣布擴大對美投資，期間亦傳出潛在晶片關稅與收購 Intel 晶圓製造部門等負面消息。

文章數: 2,547 筆

```
artDate 已成功轉換，合併後的 DataFrame :
<class 'pandas.core.frame.DataFrame'>
Index: 2547 entries, 0 to 2745
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   system_id       2547 non-null   object
1   artUrl          2547 non-null   object
2   artTitle        2547 non-null   object
3   artDate         2547 non-null   object
4   artPoster       2547 non-null   object
5   artCatagory     2546 non-null   object
6   artContent      2541 non-null   object
7   artComment      2546 non-null   object
8   e_ip            2279 non-null   object
9   insertedDate    2365 non-null   object
10  dataSource       2365 non-null   object
dtypes: object(11)
memory usage: 238.8+ KB
```

圖一、資料簡介

二、資料前處理

1. 資料清理後，首次斷詞、斷句、移除停用字之文字雲:



圖二、第一次清理之文字雲

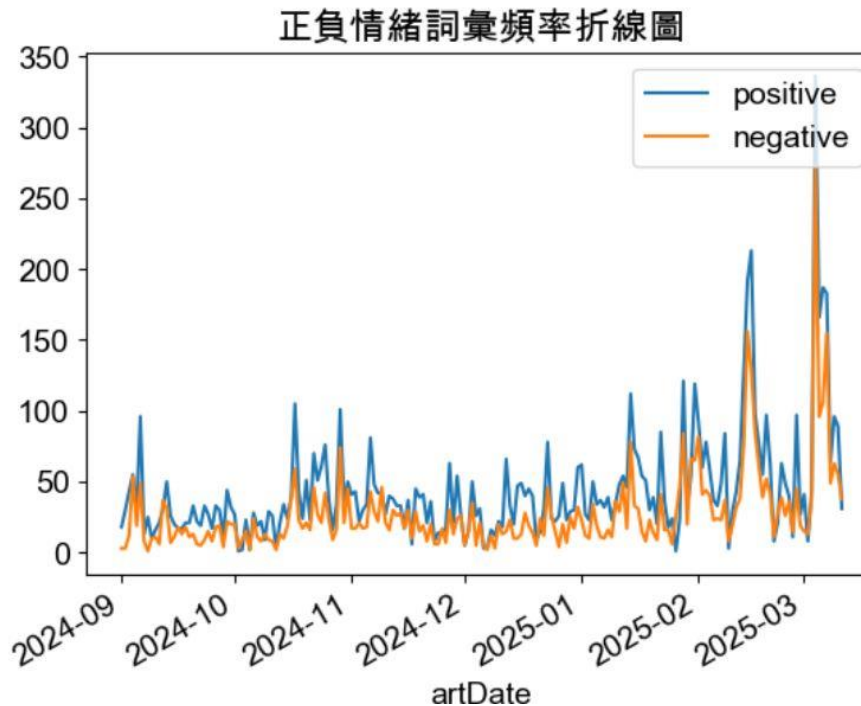
2. 經審視第一次清理之結果，手動更新停用字字典，移除「台積電」、「股市」、「報導」等與股市相關的詞彙(約 60 個)。



圖三、更新停用字後之文字雲

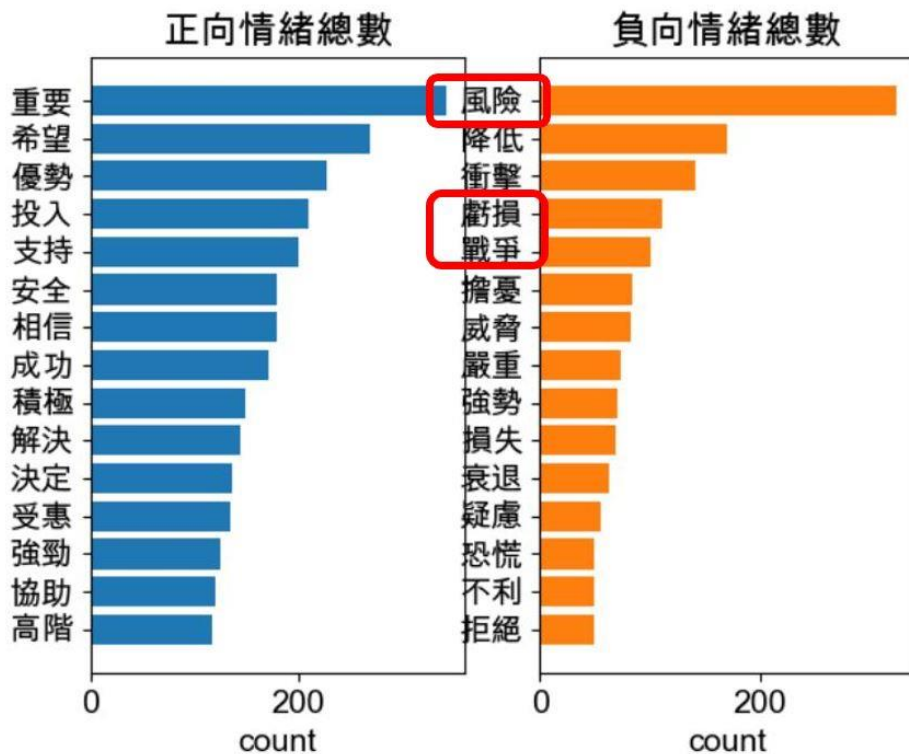
三、情緒分析(字典法)

1. 由 2024 年 9 月美國總統川普發言至 2025 農曆前台積電赴美擴大投資的傳言出現時間，由圖四觀察，無論正面或負面詞彙的頻率皆增加，代表輿論討論度變高，甚至到了 3 月初達到最高點。



圖四、正負面情緒詞統計(按月)

2. 由正向與負向情緒統計，如預期因為海外設廠成本導致的「虧損」、技術外流等「風險」、甚至是地緣政治的「戰爭」疑慮等詞彙，皆出現在統計圖中。



圖五、正負向情緒詞彙統計

3. 因此我們進一步篩選情緒種類 ['sentiment_class'] == "負向")的文章，由文字

4. 與本文一開始提及的三大因素:擴大投資、關稅議題與投資 Intel 議題,從圖六亦可觀察到「關稅」和「英特爾」都在榜上,因此我們嘗試把「關稅」加入篩選範圍,觀察其負面詞彙。



5. 利用「關稅」查詢屬於負面情緒的原文，多在探討台積電的基本面、競爭力、潛在利空和對台灣產業，甚至房地產的影響。

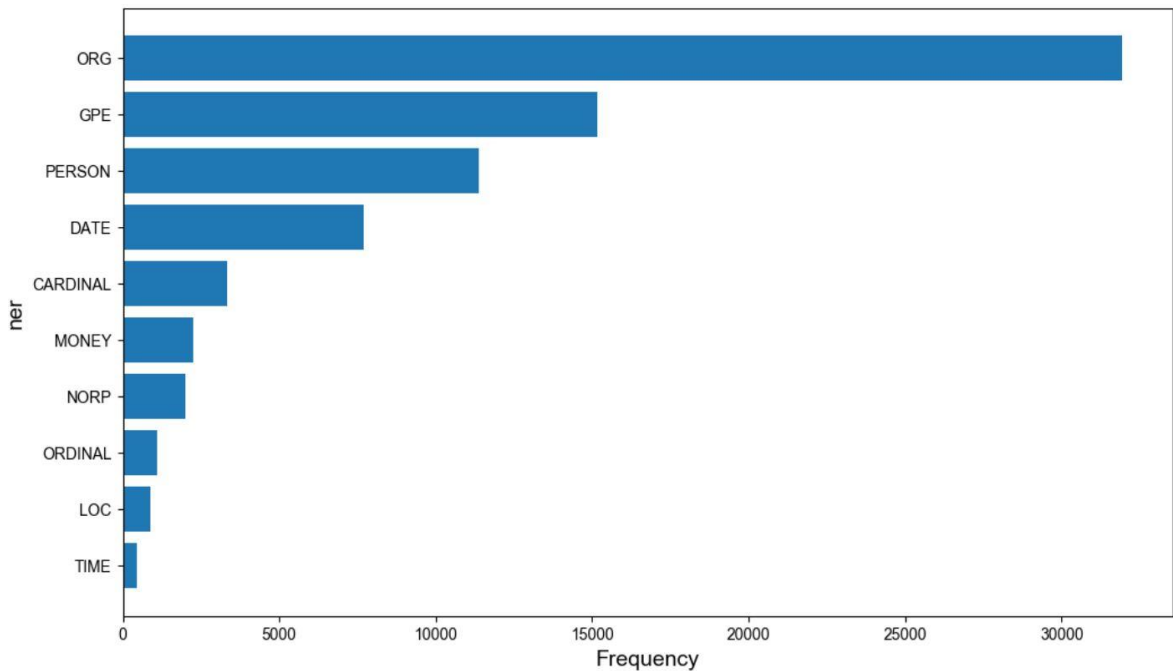
	artUrl	artTitle	artDate	artContent	sentence	sentiment_class
24725	https://www.ptt.cc/bbs/Stock/M.1729331653.A.74...	[標的]2330台積電 利空探討	2024-10-19	標的：台積電 2330 分類：討論 分析/正文：\n\n在一片樂觀之際，提出這樣...	台積電獨霸一方	負向
24726	https://www.ptt.cc/bbs/Stock/M.1729331653.A.74...	[標的]2330台積電 利空探討	2024-10-19	標的：台積電 2330 分類：討論 分析/正文：\n\n在一片樂觀之際，提出這樣...	自然是飛龍在天格局	負向
24727	https://www.ptt.cc/bbs/Stock/M.1729331653.A.74...	[標的]2330台積電 利空探討	2024-10-19	標的：台積電 2330 分類：討論 分析/正文：\n\n在一片樂觀之際，提出這樣...	但飛龍在天後	負向
24728	https://www.ptt.cc/bbs/Stock/M.1729331653.A.74...	[標的]2330台積電 利空探討	2024-10-19	標的：台積電 2330 分類：討論 分析/正文：\n\n在一片樂觀之際，提出這樣...	面對的是亢龍有悔	負向
24729	https://www.ptt.cc/bbs/Stock/M.1729331653.A.74...	[標的]2330台積電 利空探討	2024-10-19	標的：台積電 2330 分類：討論 分析/正文：\n\n在一片樂觀之際，提出這樣...	因著地緣政治	負向
...
120577	https://www.ptt.cc/bbs/Stock/M.1741594641.A.52...	[新聞]台積電加碼 對美投資，晶片 戰爭作者	2025-03-10	原文標題：台積電加碼對美投資， 晶片戰爭作者：仍根留台灣 文連結：\nhttps://...	台海安全	負向
120578	https://www.ptt.cc/bbs/Stock/M.1741594641.A.52...	[新聞]台積電加碼 對美投資，晶片 戰爭作者	2025-03-10	原文標題：台積電加碼對美投資， 晶片戰爭作者：仍根留台灣 文連結：\nhttps://...	毛利率靠漲價就可以解決 製程領先就是一口價	負向
120579	https://www.ptt.cc/bbs/Stock/M.1741594641.A.52...	[新聞]台積電加碼 對美投資，晶片 戰爭作者	2025-03-10	原文標題：台積電加碼對美投資， 晶片戰爭作者：仍根留台灣 文連結：\nhttps://...	他也提到無論如何台灣 仍然會保留全球晶片主要的產能	負向
120580	https://www.ptt.cc/bbs/Stock/M.1741594641.A.52...	[新聞]台積電加碼 對美投資，晶片 戰爭作者	2025-03-10	原文標題：台積電加碼對美投資， 晶片戰爭作者：仍根留台灣 文連結：\nhttps://...	你可以對鄉民質疑 但 你要質疑晶片戰爭作者嗎	負向
120581	https://www.ptt.cc/bbs/Stock/M.1741594641.A.52...	[新聞]台積電加碼 對美投資，晶片 戰爭作者	2025-03-10	原文標題：台積電加碼對美投資， 晶片戰爭作者：仍根留台灣 文連結：\nhttps://...	Sent from MeowPtt on my iPhone	負向

619 rows × 6 columns

圖八、由「關稅」篩選之負面情緒文章列表

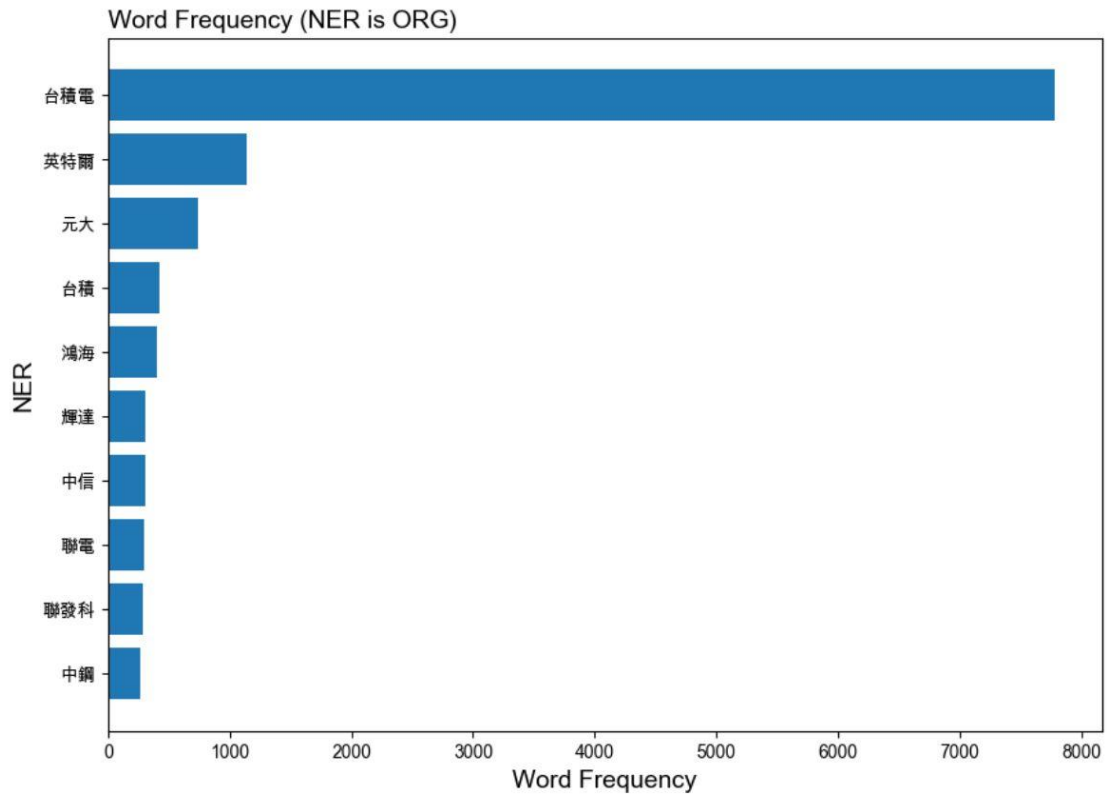
四、NER 分析(CKIP)

1. 使用 NER 分析實體種類，此份文本中，組織(ORG)出現頻率最高，第二高的實體種類為政治實體(GPE)，再來依序為人物、日期、數字。



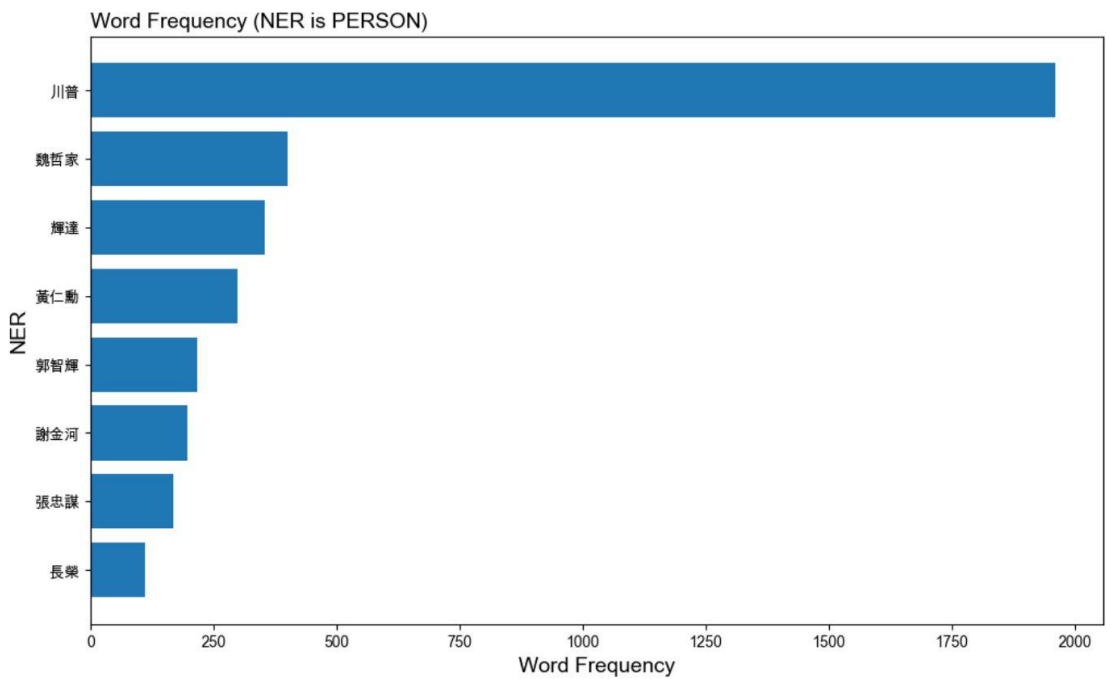
圖九、詞彙實體類別統計圖

2. 針對頻率最高的組織實體，參考圖十發現，首位為起初篩選的關鍵字「台積電」，接續則為「英特爾」、「元大」、「台灣」等等。



圖十、組織實體詞彙統計

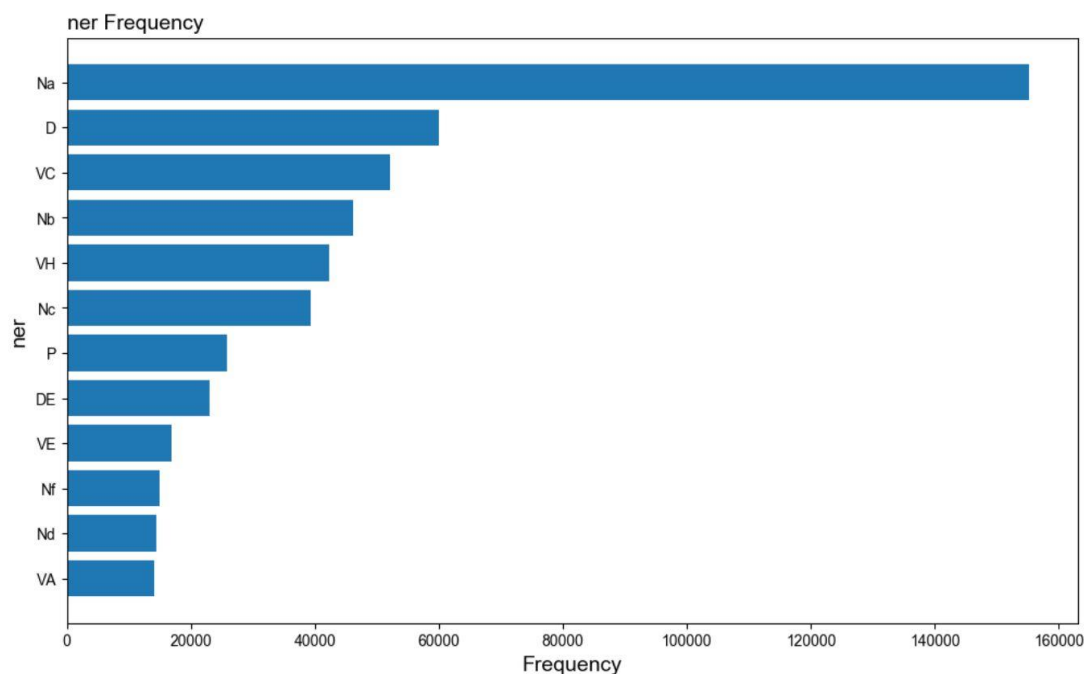
3. 第三高的實體「人物」，出現次數最高的為美國總統「川普」，其次為台積電執行長「魏哲家」，其中也包含在資料期間常在媒體發言的經濟部長「郭智輝」



圖十一、人物實體詞彙統計

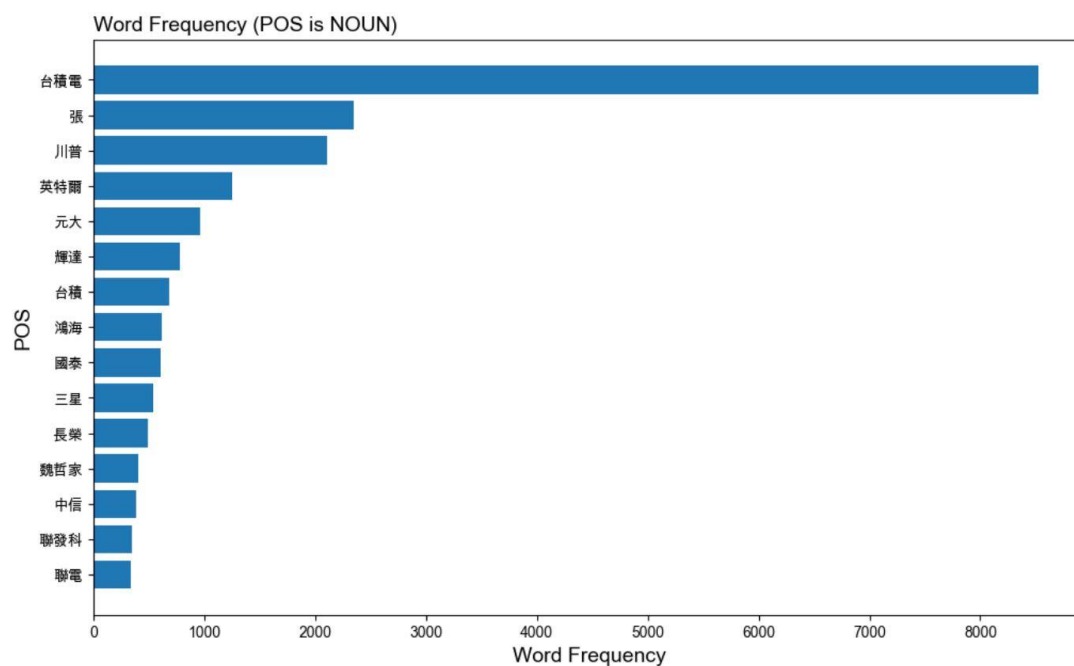
五、POS 分析

1. 由 POS 分析詞性頻率，Na 普通名詞、D 負詞和 VC 及物動詞出現最多。



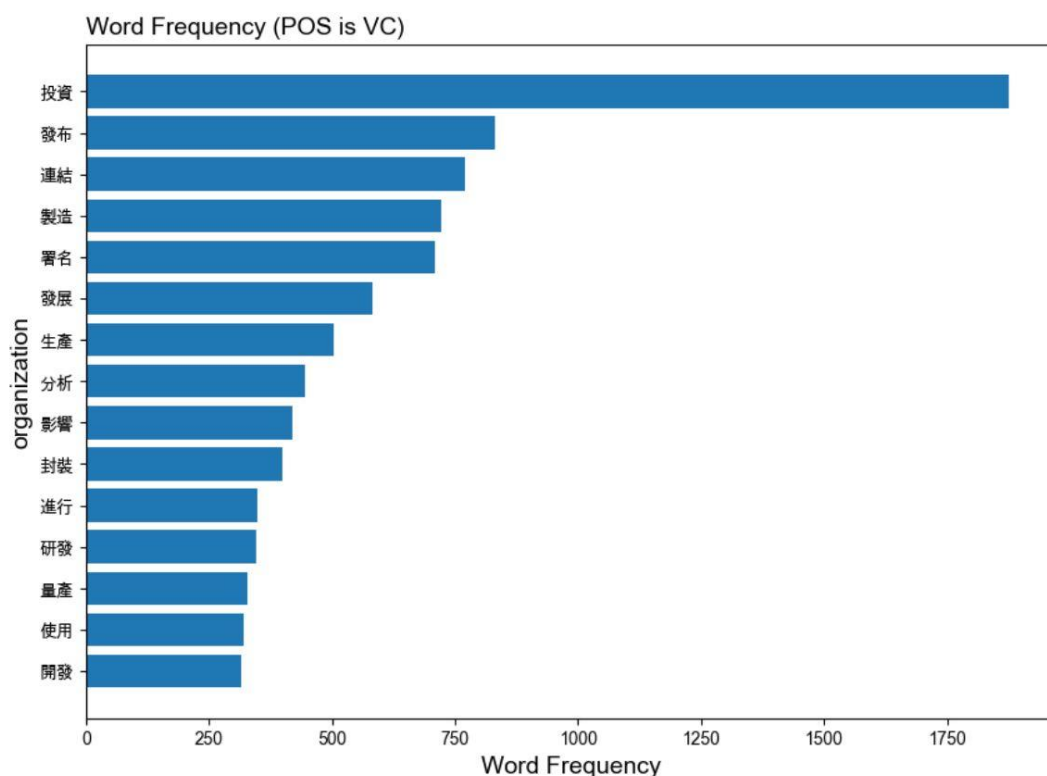
圖十二、詞性頻率統計表

2. 在普通名詞中，可以發現此日期區間，與台積電共同被探討的名詞關鍵字多為「川普」、「英特爾」、「輝達」、「三星」等，多為公司上下游單位與利益相關人士有關，和公司近期所面臨的事件相符合。



圖十三、普通名詞出現頻率表

- 針對動詞分析，我們可以找出近期談論台積電時大部分討論到的動詞，與海外投資、製造、訊息發布等相關行為。



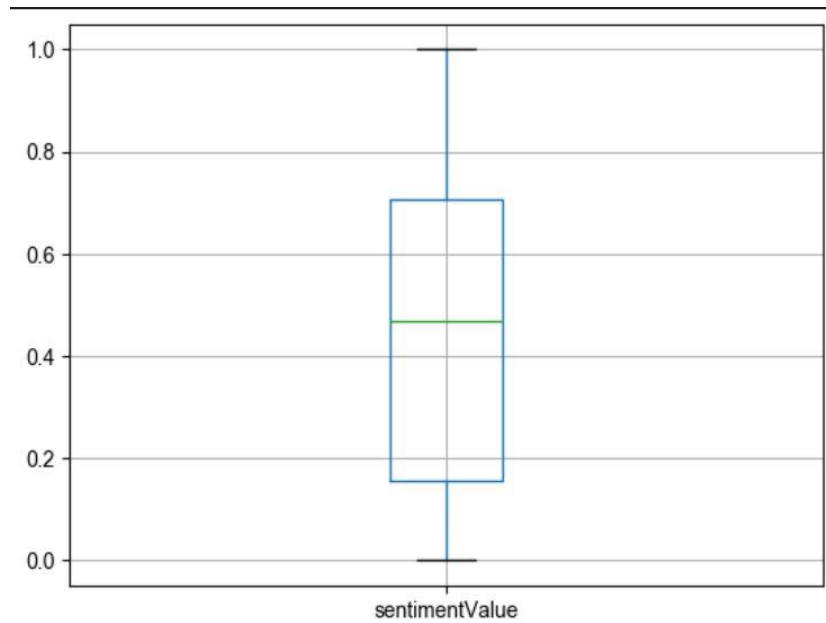
圖十四、及物動詞出現頻率表

六、情緒分析(SnowNLP)

利用 SnowNLP 進行情緒分析，情緒分數由小到大、負面至正面為 0~1。

- 情緒分布盒狀圖與數值分布:

count	mean	std	25%	50%	75%
85,630	0.44601	0.31347	0.15633	0.46758	0.70538



圖十五、情緒分布和盒狀圖

2. 從平均情緒分數趨勢觀察，大部分分數落在 0.4 至 0.5 間，且波動率不大，最低點則落在 2025 年 1 月間。



圖十六、平均情緒分數(按月)

原先本組假設，隨日期越接近台積電宣布對美擴大投資，情緒分數應會越低，但由資料情感分析後，總體情緒分數與本組一開始假設並不相同。