

Project 2_Team24 倆萬捌千捌

組員：11221828030 范瑞洋

11221828031 王上豪

A.研究主題

針對台積電在不同新聞房市板中的文字情緒分析，以及台積電、高雄與房價之間的關聯探討。

B.動機和分析目的

近幾年全台各縣市紛紛吹起「台積電旋風」，台積電在桃園、新竹、苗栗、台中、嘉義、台南與高雄等地設廠或宣布即將設廠。半導體業之高薪吸引到眾多社會新鮮人趨之若鶩，也讓不少上班族改投台積電懷抱，大量的就業人口連帶帶動當地房地產大量剛性需求，也因此導致只要有台積電設廠之處，近幾年房價也隨之水漲船高。

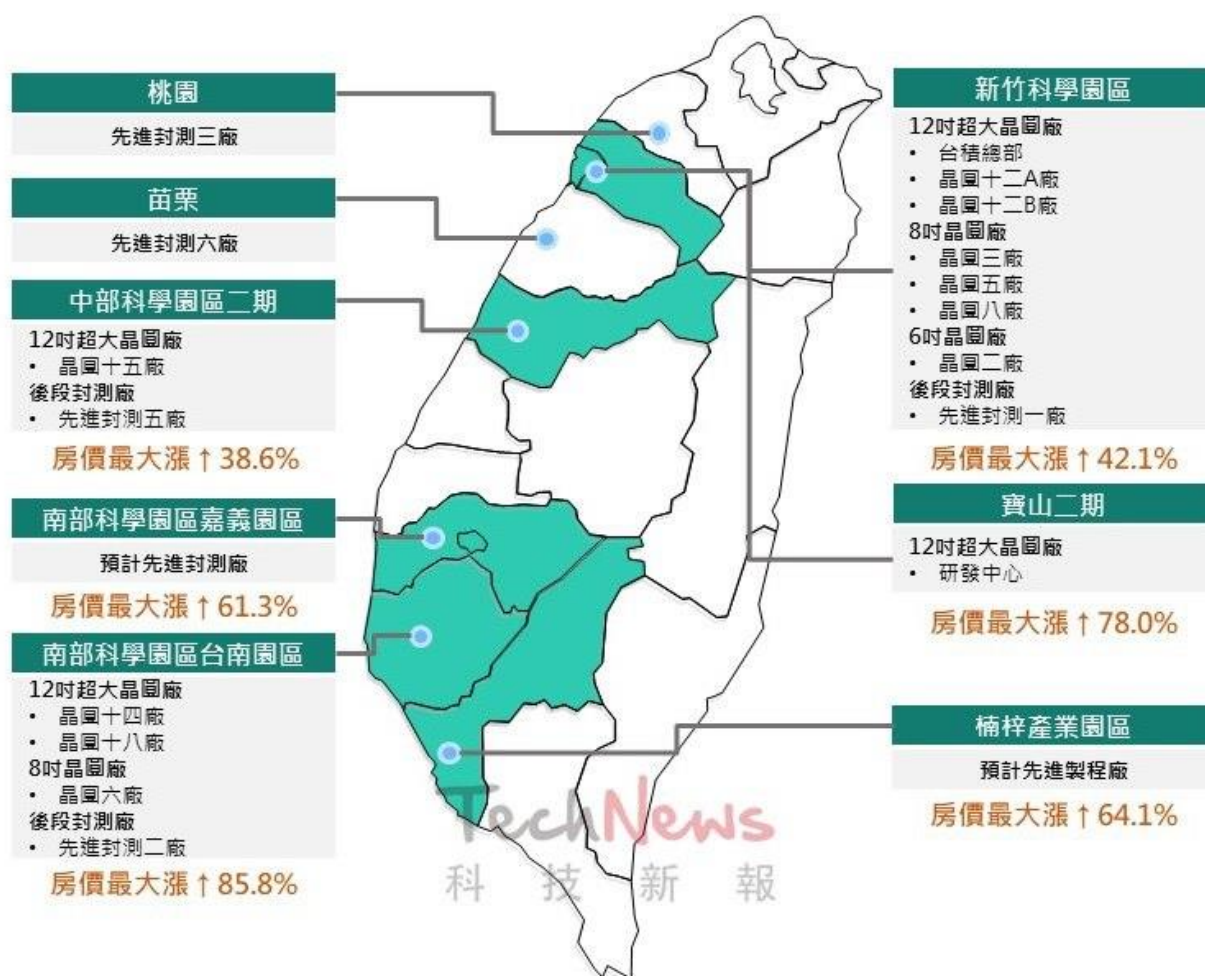
依據住商不動產的資料顯示，從 2019 年至 2023 年短短 4 年內，因「台積電效應」導致周遭房市暴漲，如新竹科學園區房價最大漲幅達 42.1%、寶山二期最大漲幅達 78.0%、台中科學園區最大漲幅達 38.6%、南部科學園區台南園區房價最大漲幅達 85.8%，連尚未蓋好工廠的南部科學園區嘉義園區、楠梓產業園區房價最大漲幅也達 61.3%、64.1%，顯現只要是台積電所到之處房價皆有相當程度的漲幅。

因此，本研究欲透過台灣幾個重要新聞媒體：自由時報、東森新聞與聯合報，了解台灣近一年半以來，台積電在房市新聞報導下，網絡圖之關係。此外，不同新聞媒體通常背後皆有代表不同的政黨色彩，因此也愈透過情緒分析來了解每日文章係偏向正面抑或是為負面報導，以了解是否會因為不同

報章雜誌，會在同一報導基礎上，而有不同的報導取向。

產業園區	影響行政區	2019年平均單價 (萬元/坪)	2023年平均單價 (萬元/坪)	漲幅
新竹科學園區寶山用地擴建二期	新竹縣寶山鄉	13.88	24.7	77.95%
新竹科學園區寶山用地擴建二期	新竹市	22.76	32.35	42.14%
中部科學園區台中園區擴建二期	台中市大雅區	18.64	24.12	29.40%
中部科學園區台中園區擴建二期	台中市沙鹿區	17.4	24.12	38.62%
中部科學園區台中園區擴建二期	台中市西屯區	25.52	31.96	25.24%
南部科學園區嘉義園區	嘉義縣太保市	9.18	14.81	61.33%
南部科學園區嘉義園區	嘉義縣朴子市	8.47	13.64	61.04%
南部科學園區台南園區	台南市善化區	15.25	28.33	85.77%
南部科學園區台南園區	台南市新市區	15.01	27.64	84.14%
南部科學園區台南園區	台南市安定區	14.51	21.85	50.59%
楠梓產業園區	高雄市楠梓區	18.35	23.78	29.59%
楠梓產業園區	高雄市橋頭區	15.11	24.79	64.06%
楠梓產業園區	高雄市仁武區	15.92	25.74	61.68%

「台積電效應」下的房市暴漲



C. 資料集描述

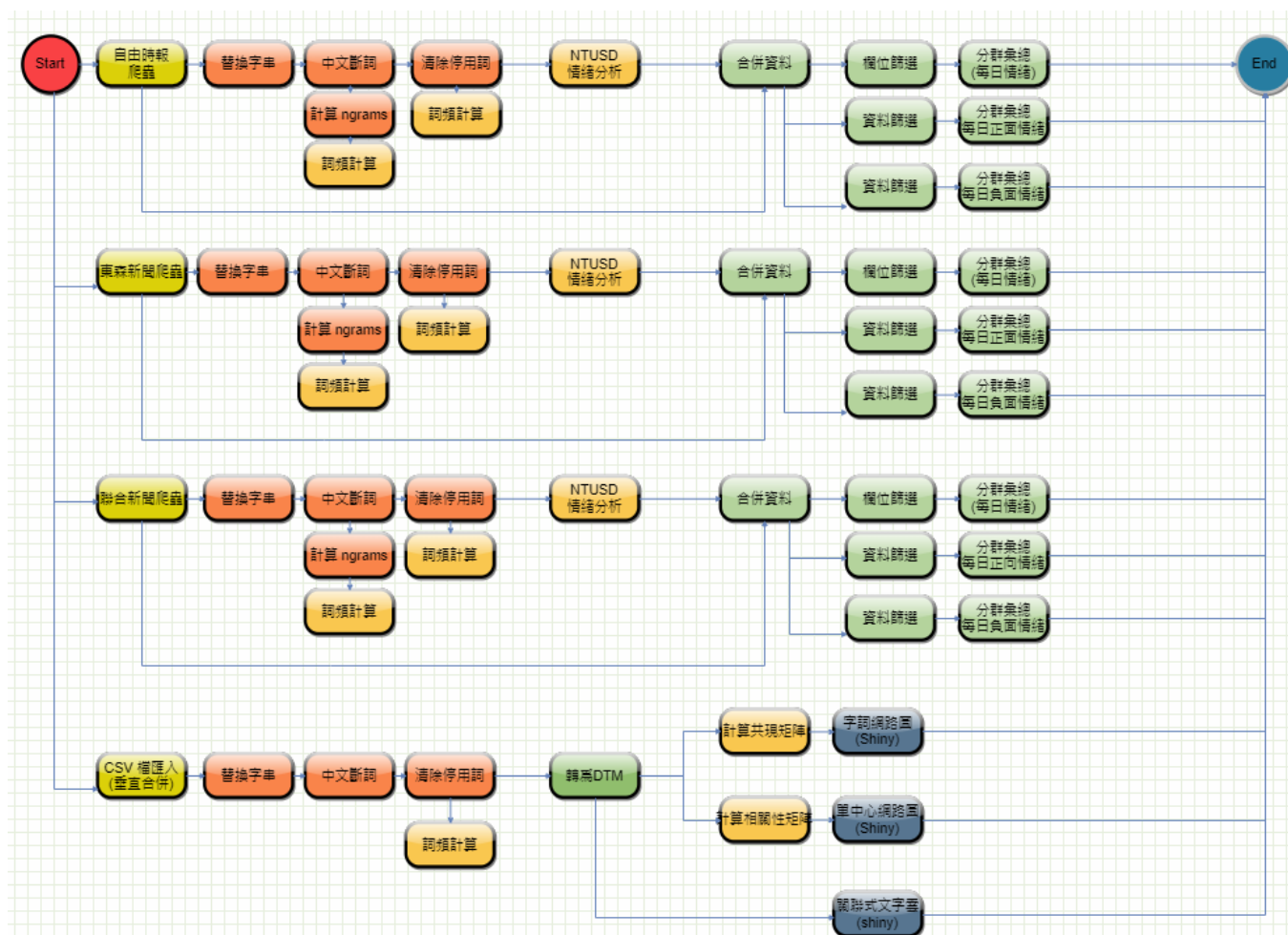
1. 資料來源：

- (1) 自由時報 - 地產版
- (2) 東森新聞 - 房產版
- (3) 聯合報 - 房市版 (因 Tarflow 上無房市版，因此以自行爬蟲方式取得資料)

2. 搜尋關鍵字：台積電

3. 資料蒐集時間：2022/12/01 ~ 2024/05/19

D. 資料分析步驟



1. 匯入資料(透過爬蟲抓取自由時報、東森新聞、聯合報資料)
2. 資料預處理(替換字串、中文斷詞、清除停用字等)
3. 進行情緒分析(NTUSD)
4. 合併資料(將情緒分析後資料與原始資料合併)
5. 分析每日情緒(正面、負面、彙整)
6. 垂直合併(自由時報+東森新聞+聯合報)
7. 資料預處理(替換字串、中文斷詞、清除停用字等)
8. 轉換成 DTM
9. 視覺化呈現(關聯性文字雲、字詞網絡圖、單中心網絡圖)

E. 資料分析內容與說明

1. 匯入資料

(1) 自由時報爬蟲

以「台積電」為關鍵字，抓取【自由時報 - 地產版】近一年半的資料 (2022/12/01 ~ 2024/05/19)，共抓取 274 筆資料。

自由時報 爬蟲 (4)

The screenshot displays the 'Free Press Crawler' (自由時報 爬蟲) interface. It features a 'Parameter Settings' (參數設定) tab and a 'Task Results' (任務結果) tab. Under 'Parameter Settings', there is a 'Selection Board' (選擇看板) with a list of categories: Yunlin (雲林縣), art (藝文), auto (汽車), de (地產), ec (財經), ent (娛樂), and estate (地產). The 'Search Keywords' (搜尋關鍵字) field is empty. The 'Exclude Keywords' (排除關鍵字) field contains the text '以換行區隔, e.g. 雲山動物園 獅子'. The 'Search Start Date' (搜尋起始日期) is set to 2022/12/01, and the 'Search End Date' (搜尋結束日期) is set to 2024/05/19. A green bar at the bottom indicates 'Save Settings' (儲存設定).

(2) 東森新聞爬蟲

以「台積電」為關鍵字，抓取【東森新聞 - 房產版】近一年半的資料(2022/12/01 ~ 2024/05/19)，共抓取 90 筆資料。

東森新聞爬蟲 (24)

參數設定	任務結果
<div>選擇看板 *</div> <div>business(財經) car(汽車) china(兩岸) entertainment(娛樂) fun(新奇) health(健康) house(房產)</div> <div>搜尋關鍵字</div> <div>台積電</div> <div>排除關鍵字</div> <div>以換行區隔，e.g. 泰山動物園 獅子 —</div> <div>搜尋起始日期</div> <div>2022/12/01</div> <div>搜尋結束日期</div> <div>2024/05/19</div> <div>儲存更改</div>	

(3) 聯合報爬蟲

以「台積電」為關鍵字，透過自行爬蟲的方式，抓取【聯合報 - 房產版】近一年半的資料(2022/12/01 ~ 2024/05/19)後，將資料進行整理並儲存成 CSV 檔後匯入，共抓取 589 筆資料。

聯合新聞爬蟲 (102)

參數設定	任務結果
<div>選擇文件 *</div> <div>0523.csv</div> <div>是否使用header *</div> <div>是</div> <div>儲存更改</div>	

2. 資料預處理(替換字串、中文斷詞、清除停用字等)

(1) 自由時報

- 替換字串

替換字串 (30)

參數設定	Input - 4	任務結果
<div>選擇處理欄位 *</div> <div>artContent</div> <div>選擇替換規則檔案</div> <div>請選擇</div> <div>替換字串設定</div> <div>以換行符號區隔，e.g. 被替換的字詞>>替換後的字詞 全聯福利中心>>全聯 臺電>>台電</div> <div>儲存更改</div>		

- 中文斷詞

透過 ngrams($n = 2$)，以詞頻計算方式，觀察同時出現且頻率較高的詞彙，如楠梓產業園區、仁武產業園區、北高雄、先進製程、高鐵左營站、漢神巨蛋、實價登錄、半導體產業、S 廊帶、高大特區、橋頭科學園區、高雄大學、橋頭新市鎮等，再把不小心被斷詞斷開的詞彙加回去。

中文斷詞 (31)

- 清除停用字

透過詞頻計算，觀察不需要的字，如成為、目前、今年、未來、預計、表示、去年、包括、陸續、指出等，並加入字定義停止詞中。

清除停用詞 (32)

(2) 東森新聞

- 替換字串

≡ 替換字串 (89)

參數設定

Input - 24

任務結果

選擇處理欄位 *

artContent

選擇替換規則檔案 ⓘ

請選擇.....

替換字串設定 ⓘ

以換行符號區隔，e.g.
被替換的字詞>>替換後的字詞
全聯福利中心>>全聯
藍電>>台灣

儲存更改

● 中文斷詞

透過 ngrams($n = 2$)，以詞頻計算方式，觀察同時出現且頻率較高的詞彙，如科技之心、高雄市長、馬稠後產業園區、新竹科學園區、楠梓園區、新青安、陽明交大、產業轉型等，再把不小心被斷詞斷開的詞彙加回去。

≡ 中文斷詞 (31)

參數設定

Input - 30

任務結果

選擇處理欄位 *

result

選擇字典 ⓘ

請選擇.....

定義詞彙 ⓘ

楊得基新聞區 100
仁武舊新聞區 100
尖高路 100
先進制程 100
高雄左營站 100

儲存更改

● 清除停用字

透過詞頻計算，觀察不需要的字，如成為、目前、今年、持續、未來、預計、表示、去年、提供、包括、陸續、指出等，並加入字定義停止詞中。

≡ 清除停用詞 (91)

參數設定

Input - 90

任務結果

語言 *

Chinese

是否清除單字元 ⓘ

是

清除英文字母 *

否

清除換行符號 *

是

清除html tag *

是

使用預設停止詞

是

是否轉為小寫英文

是

清除數字 *

是

清除特殊標點符號 *

是

自定義停止詞

萬元
成為
目前
今年
持續

儲存更改

(3) 聯合報

- 替換字串

≡ 替換字串 (104)

參數設定 Input - 102 任務結果

選擇處理欄位 *
artContent

選擇替換規則檔案 ⓘ
請選擇.....

替換字串設定 ⓘ
以換行符號區隔，e.g. 被替換的字詞>>替換後的字詞
金聯福利中心>>金聯
臺灣>>台灣

儲存更改

- 中文斷詞

透過 ngrams($n = 2$)，以詞頻計算方式，觀察同時出現且頻率較高的詞彙，如科技之心、高雄市長、1.4 奈米、2 奈米、3 奈米、新竹科學園區、南台灣、中油高煉廠、產業園區、景觀大樓、南門圓環、新竹縣市等，再把不小心被斷詞斷開的詞彙加回去。

≡ 中文斷詞 (106)

參數設定 Input - 104 任務結果

選擇處理欄位 *
result

選取字典 ⓘ
請選擇.....

定義詞彙 ⓘ
科技之心 100
高雄市長 100
鹿港後港產業區 100
1.4奈米 100
2奈米 100

儲存更改

- 清除停用字

透過詞頻計算，觀察不需要的字，如今年、表示、目前、未來、去年、近年、指出、成為、包括、位於、認為、預計等，並加入字定義停止詞中。

≡ 清除停用詞 (108)

參數設定 Input - 106 任務結果

語言 *
Chinese

是否清除單字元 ⓘ
是

清除英文字母 *
否

清除換行符號 *
是

清除html tag *
是

使用預設停止詞
是

是否轉為小寫英文
是

清除數字 *
是

清除特殊標點符號 *
是

自定義停止詞
今年
表示
目前
未來
去年

儲存更改

3. 進行情緒分析(NTUSD)

(1) 自由時報

選用 NTUSD 情緒分析，並使用否定字及加強字進行分析，每筆資料中最大正面情緒為 37、最小正面情緒為 2；最大負面情緒為 19、最小負面情緒為 0。此外，從情緒貢獻長條圖中可以發現，正向情緒中前 15 名常出現字及出現頻率為：發展(199)、建設(148)、開發(142)、吸引(129)、積極(102)、增加(96)、成長(89)、優勢(84)、重要(66)、機會(63)、擁有(63)、進行(61)、提高(60)、明顯(60)、便利(55)；負向情緒中前 15 名常出現字及出現頻率為：超過(78)、可望(49)、減少(38)、問題(32)、不斷(32)、揭露(30)、壓力(28)、刺激(28)、驚人(27)、負擔(26)、降低(20)、高價(20)、衝擊(20)、不到(19)、干擾(19)。

NTUSD 情緒分析 (53)

Parameter Setting | Input - 32 | Task Result

Whether to use negation words: ☒ Yes

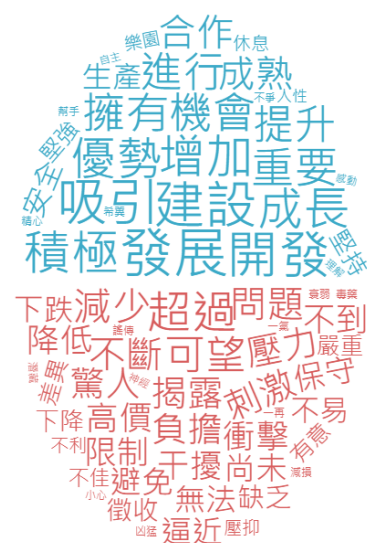
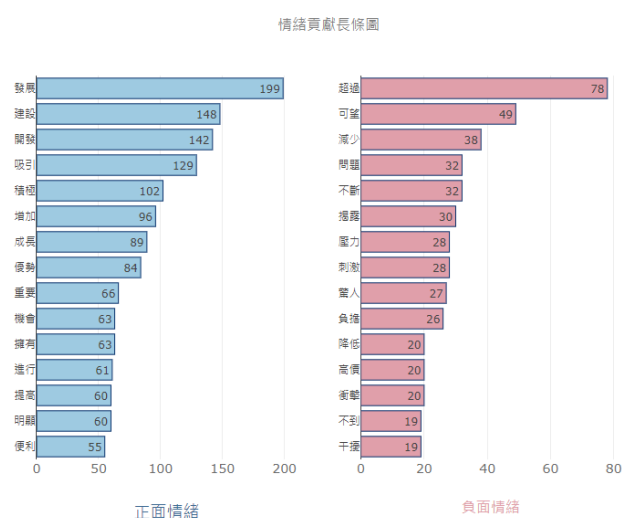
Whether to use emphasis words: ☒ Yes

Define positive word list:
以換行符號區隔，e.g.
高興
快樂
開心...

Define negative word list:
以換行符號區隔，e.g.
難過
生氣
恐懼...

Remove emotion words:
以換行符號區隔，e.g.
難過
高興
恐懼...

Save Settings



(2) 東森新聞

選用 NTUSD 情緒分析，並使用否定字及加強字進行分析，每筆資料中最大正面情緒為 32、最小正面情緒為 3；最大負面情緒為 32、最小負面情緒為 0。此外，從情緒貢獻長條圖中可以發現，正向情緒中前 15 名常出現字及出現頻率為：建設(43)、發展(38)、增加(36)、機會(31)、開發(29)、成長(20)、吸引(20)、確實(20)、明顯(20)、穩定(18)、專家(17)、優勢(17)、積極(17)、擁有(17)、希望(14)；負向情緒中前 15 名常出現字及出現頻率為：問題(22)、不斷(18)、超過(16)、無法(15)、透露(14)、減少(11)、驚人(10)、衝擊(10)、擔心(9)、負擔(9)、嚴重(8)、保守(8)、限制(8)、逼近(8)、禁止(7)。

NTUSD 情緒分析 (124)

參數設定

Input - 91

任務結果

是否使用否定詞

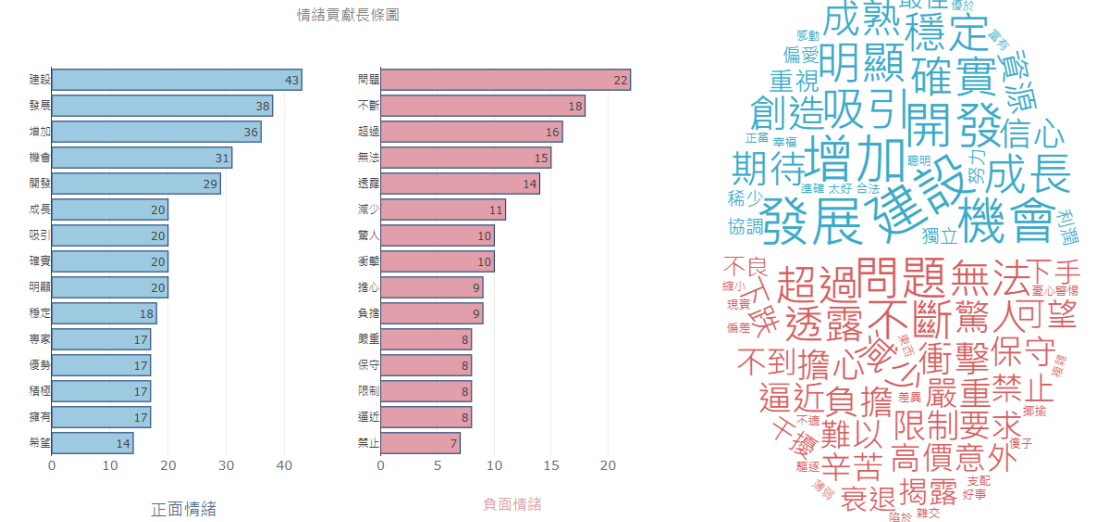
是否使用加強詞

定義正面詞彙

定義負面詞彙

移除情緒詞

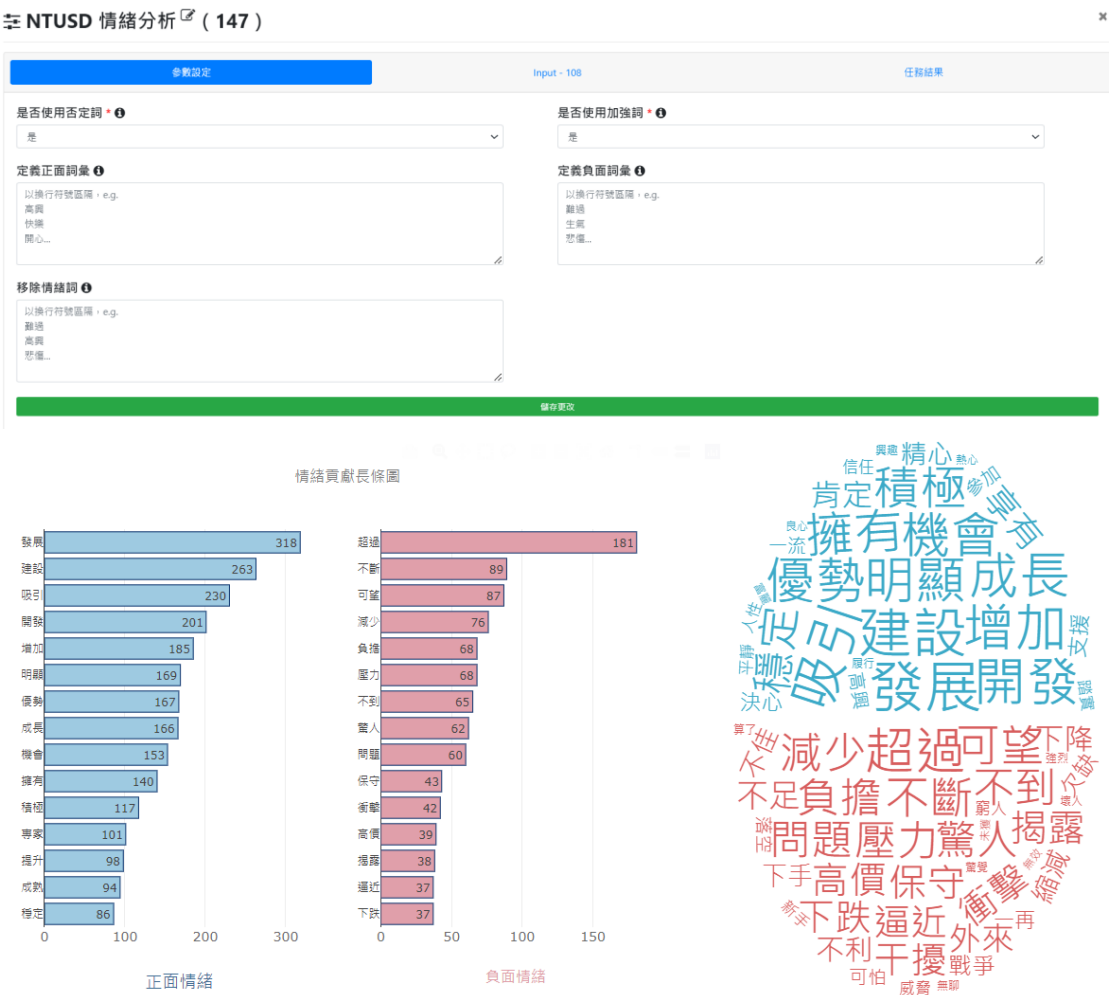
儲存更改



(3) 聯合報

選用 NTUSD 情緒分析，並使用否定字及加強字進行分析，每筆資料中最大正面情緒為 44、最小正面情緒為 1；最大負面情緒為 41、最

小負面情緒為 0。此外，從情緒貢獻長條圖中可以發現，正向情緒中前 15 名常出現字及出現頻率為：發展(318)、建設(263)、吸引(230)、開發(201)、增加(185)、明顯(169)、優勢(167)、成長(166)、機會(153)、擁有(140)、積極(117)、專家(101)、提升(98)、成熟(94)、穩定(86)；負向情緒中前 15 名常出現字及出現頻率為：超過(181)、不斷(89)、可望(87)、減少(76)、負擔(68)、壓力(68)、不到(65)、驚人(62)、問題(60)、保守(43)、衝擊(42)、高價(39)、揭露(38)、逼近(37)、下跌(37)。



4. 合併資料

(1) 自由時報

將情緒分析完的資料與原始資料進行合併。

合併資料 (64)

參數設定 Input - 4 Input - 53 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

(2) 東森新聞

將情緒分析完的資料與原始資料進行合併。

合併資料 (134)

參數設定 Input - 24 Input - 124 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

(3) 聯合報

將情緒分析完的資料與原始資料進行合併。

合併資料 (146)

參數設定 Input - 102 Input - 147 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

5. 分析每日情緒

(1) 自由時報

● 每日情緒彙整

先將資料欄位進行篩選，僅保留日期、正面情緒分數、負面情緒分數及該篇新聞正負面情緒彙整分數(sentiment_value)，再以「日」為單位，將該日情緒彙整分數進行加總，計算每天情緒是正面抑或是負面。

參數設定

Input - 68

任務結果

選擇日期欄位 *

artDate

選擇日期類型 *

日

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

system_id

artDate

positive_count

negative_count

sentiment_value

日期格式 *

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

匯總函數 *

sum

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

system_id

artDate

positive_count

negative_count

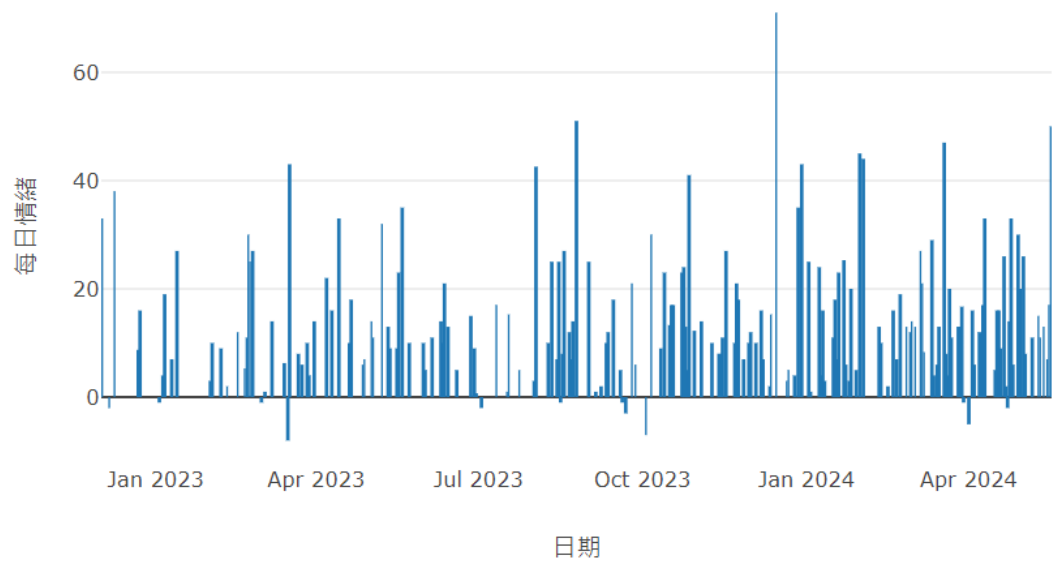
sentiment_value

儲存更改

在近一年半的期間中，共有 194 天有相關報導，其中多數日期中多為正向情緒，僅有 8 天是負面情緒 (2022/12/06、2023/01/03、2023/03/01、2023/03/16、2023/07/03、2023/08/16、2023/09/20、2023/09/22、2023/10/03、2024/03/29、2024/04/01、2024/04/23)，此 8 日當天皆僅有一篇文章，報導面相多以台積電效應降溫(某段期間曾有台積電楠梓廠房暫緩興建消息)、政策影響(平均地權)條例、打房政策)、缺工物料漲、兩岸衝突有關。以 2023/03/16 為例，當天僅有一篇新聞報導，標題為「缺工、營建成本高漲快凍未條 高雄建商『策略性停工』將達共識」，內容主要敘述因缺工、建材成本持續上升，且台積電效應當時逐步降溫下，整體高雄建商決議暫緩興建，並透過群體力量共同給予材料廠商，希望可以降價以提升利潤。

而在這些日期中，每日情緒彙總分數最高為 2023/12/15，因為當天有 3 則報導：「台積電 15 分車程 高大特區吸科技新貴」、「【高雄】美術館核心區生活圈 綠意環境建設加分」、「【台中】中科后里園區 3 字頭科技新綠洲」，此三篇報導皆與台積電在高雄與台中未來即將設廠有關，報導皆屬正面消息，因此讓該日情緒分數較高！

【自由時報】每日情緒彙總



● 每日正面情緒

先將資料欄位進行篩選，透過「`sentiment_value > 0`」條件式，篩選出情緒為正的資料，再以「日」為單位，將該日情緒彙整分數進行加總。在近一年半的期間中，有 181 天出現正面情緒報導，正面情緒分數如下圖所示。

分群彙總 (每日正面情緒) (85)

參數設定

Input - 174

任務結果

選擇日期欄位 *

artDate

選擇日期類型 *

日

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

negative_count

negative_words

booster_count

booster_words

negator_count

negator_words

sentiment_value

日期格式 *

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

匯總函數 *

sum

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

system_id

dataSource

artUrl

artDate

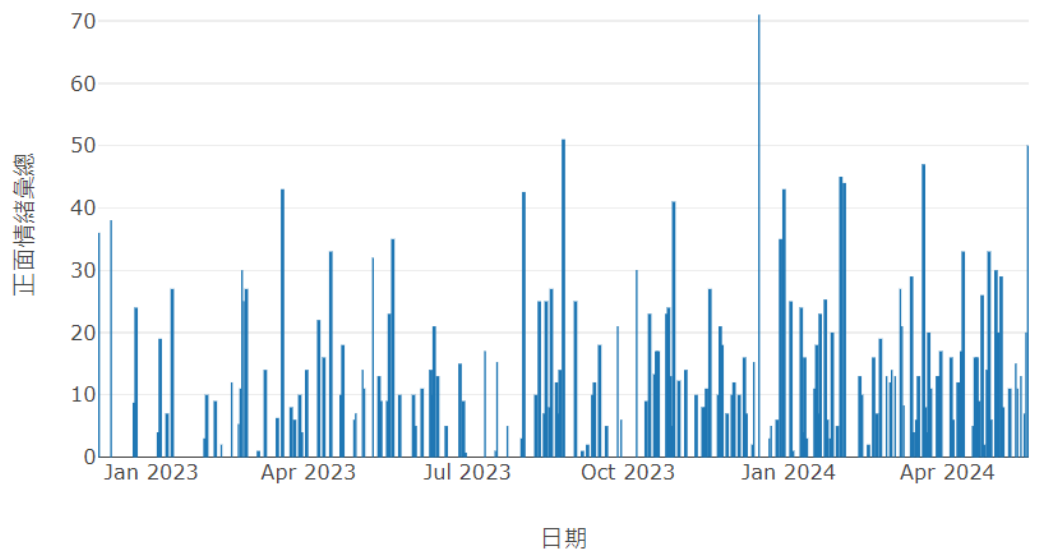
artTitle

artContent

sub_media

儲存更改

【自由時報】每日正面情緒彙總



● 每日負面情緒

先將資料欄位進行篩選，透過「`sentiment_value < 0`」條件式，篩選出情緒為負的資料，再以「日」為單位將該日情緒彙整分數進行加總。在近一年半的期間中，僅有 18 天出現負面情緒報導。

分群彙總 (每日負面情緒) (70)

參數設定

Input - 177

任務結果

選擇日期欄位 *

artDate

選擇日期類型 *

日

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

negative_count

negative_words

booster_count

booster_words

negator_count

negator_words

sentiment_value

日期格式 *

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

匯總函數 *

sum

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

system_id

dataSource

artUrl

artDate

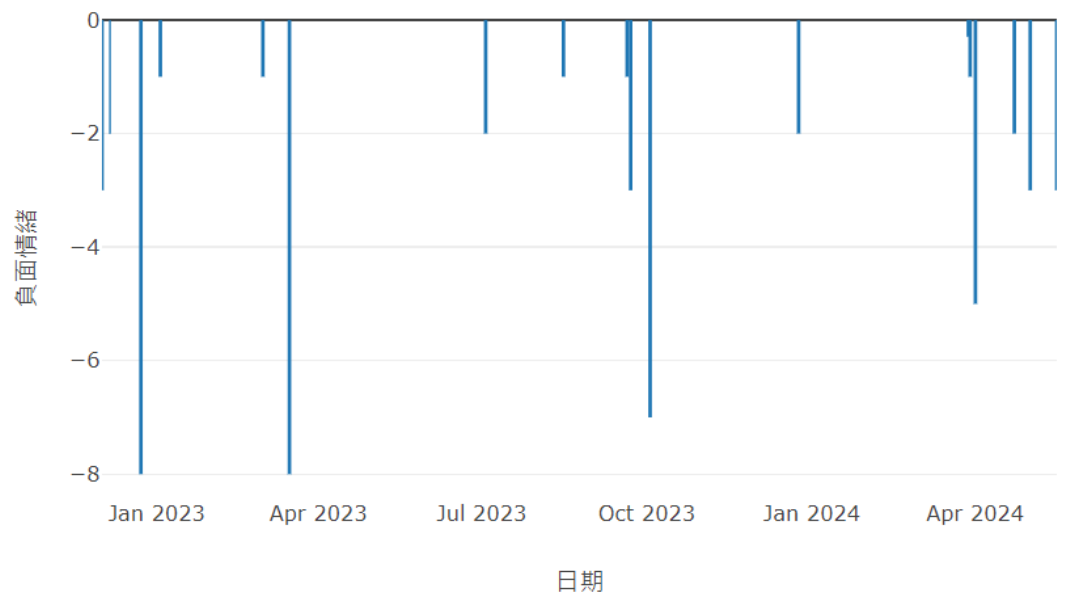
artTitle

artContent

sub_media

儲存更改

【自由時報】每日負面情緒彙總



(2) 東森新聞

● 每日情緒彙整

先將資料欄位進行篩選，僅保留日期、正面情緒分數、負面情緒分數及該篇新聞正負面情緒彙整分數(sentiment_value)，再以「日」為單位，將該日情緒彙整分數進行加總，計算每天情緒是正面抑或是負面。

分群彙總 (每日情緒) (155)

選擇日期欄位 *

antDate

選擇日期類型 *

日

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

system_id

antDate

positive_count

negative_count

sentiment_value

日期格式 *

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

匯總函數 *

sum

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

system_id

antDate

positive_count

negative_count

sentiment_value

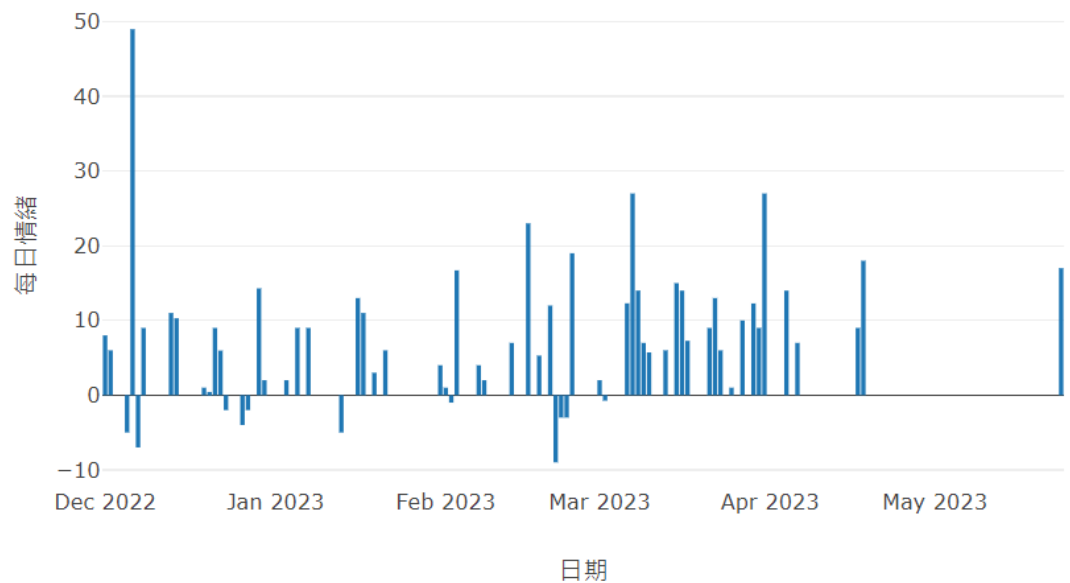
儲存設定

在近一年半的期間中，共有 64 天有相關報導，其中多數日期中多為正向情緒，僅有 17 天是負面情緒 (2022/12/05、2022/12/07、2022/12/22、2022/12/23、2022/12/26、2022/12/27、2022/12/30、2023/01/13、2023/01/19、2023/02/02、2023/02/03、2023/02/21、2023/02/22、2023/02/23、2023/03/01、2023/03/02、2023/03/06)，此 17

日當天皆僅有一篇文章，報導面相多以網友評論、建商說法、政策(打房、平均地權)、凶宅、兩岸衝突等內容有關。

而在這些日期中，每日情緒彙總分數最高為 2022/12/06，因為當天有 3 則報導：「台積電 2 奈米落腳中科新進展 專家讚：周邊房價直逼 7 期」、「新竹縣首座 Fitwel 訂製宅 落腳竹科概念區」、「台積電村開箱！攜家帶眷 4 人 全額 11 萬補助租房」，此三篇報導與台積電在新竹、台中、美國亞利桑那州廠房有關，報導皆屬正面消息，因此讓該日情緒分數較高！

【東森新聞】每日情緒彙總



- 每日正面情緒

先將資料欄位進行篩選，透過「`sentiment_value > 0`」條件式，篩選出情緒為正的資料，再以「日」為單位，將該日情緒彙整分數進行加總。在近一年半的期間中，有 53 天出現正面情緒報導，正面情緒分數如下圖所示。

分群彙總 (每日正面情緒) (158)

參數設定

Input - 180

任務結果

選擇日期欄位 *

artDate

選擇日期類型 *

日

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

negative_count

negative_words

booster_count

booster_words

negator_count

negator_words

sentiment_value

日期格式 *

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

匯總函數 *

sum

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

system_id

artTitle

artUrl

artDate

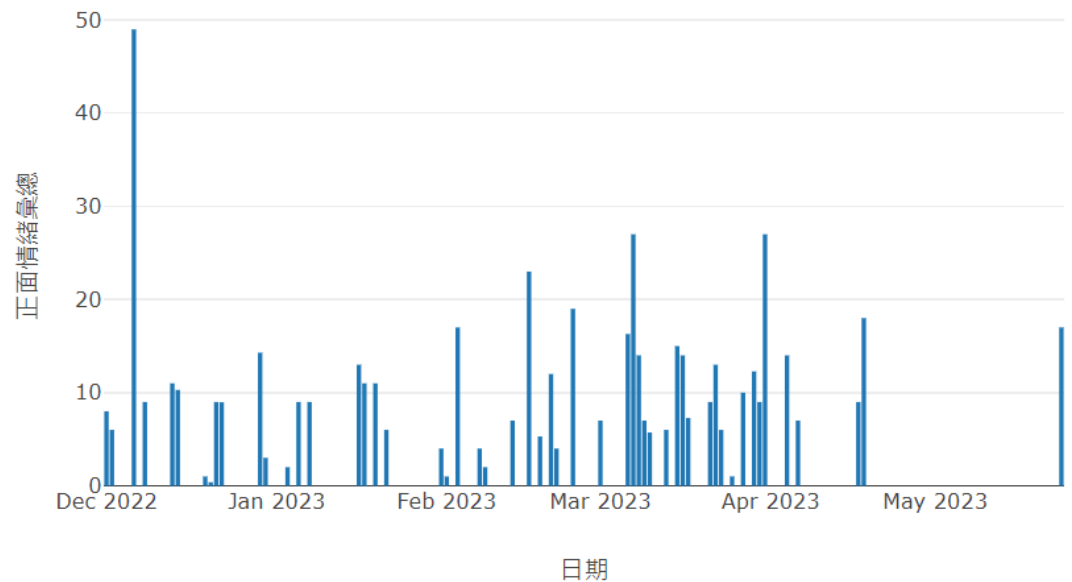
artCategory

artContent

dataSource

儲存更改

【東森新聞】每日正面情緒彙總



● 每日負面情緒

先將資料欄位進行篩選，透過「`sentiment_value < 0`」條件式，篩選出情緒為負的資料，再以「日」為單位，將該日正面情緒文章數量進行計算。將該日情緒彙整分數進行加總。在近一年半的期間中，有 17 天出現負面情緒報導，負面情緒分數如下圖所示。

分群彙總 (每日負面情緒) (159)

參數設定

Input - 181

任務結果

選擇日期欄位 *

artDate

選擇日期類型 *

日

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

negative_count

negative_words

booster_count

booster_words

negator_count

negator_words

sentiment_value

日期格式 *

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

匯總函數 *

sum

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

system_id

artTitle

artUrl

artDate

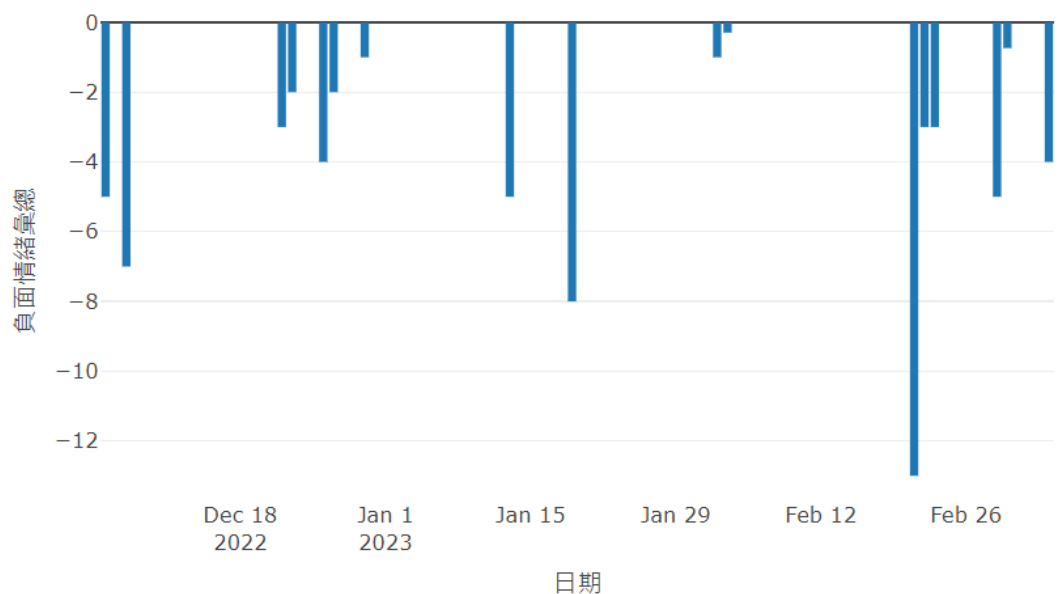
artCategory

artContent

dataSource

儲存更改

【東森新聞】每日負面情緒彙總



(3) 聯合報

● 每日情緒彙整

先將資料欄位進行篩選，僅保留日期、正面情緒分數、負面情緒分數及該篇新聞正負面情緒彙整分數(sentiment_value)，再以「日」為單位，將該日情緒彙整分數進行加總，計算每天情緒是正面抑或是負面。

參數設定

Input - 150

任務結果

選擇日期欄位 *

artDate

選擇日期類型 *

日

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

system_id
artDate
positive_count
negative_count
sentiment_value

日期格式 *

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

運總函數 *

sum

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

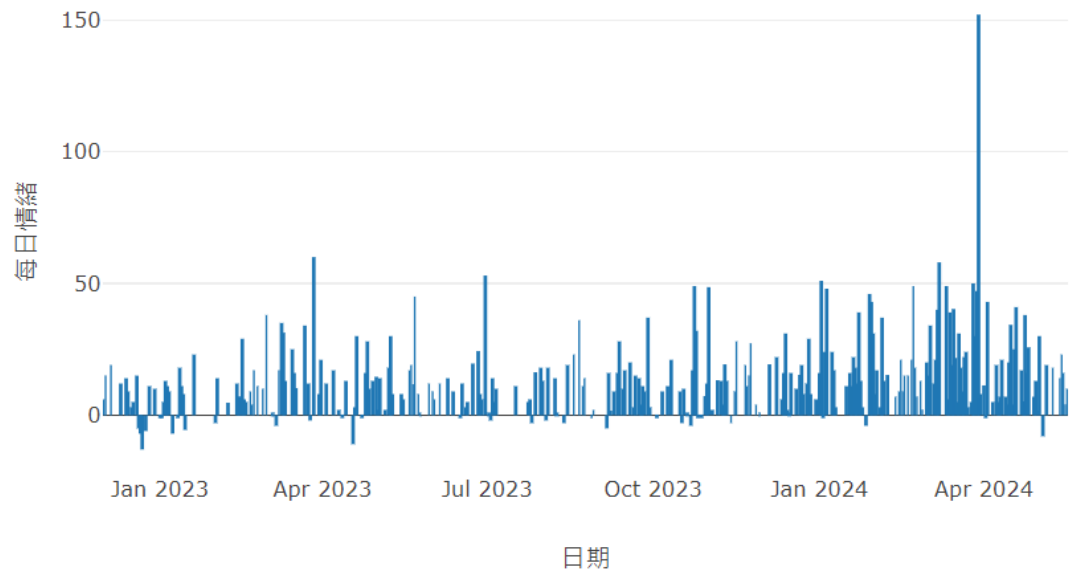
system_id
artDate
positive_count
negative_count
sentiment_value

儲存更改

在近一年半的期間中，共有 294 天有相關報導，其中多數日期中多為正向情緒，僅有 59 天是負面情緒，報導面相多以台積電建廠計畫生變、網友說法、建商說法、兩岸衝突、政策影響(打房、平均地權、屯房稅)、房市冷、青年房貸、總統大選等內容有關。而在這些日期中，

每日情緒彙總分數最高為 2024/03/29，因為當天有 12 則報導：「國際品牌亞昕集團 再創北區中國醫生活圈 28 層地標建築」、「下波房市利基？李同榮：高鐵經濟『見軌則發』」、「高雄觀光發威 百億商機爆發 商圈店面翻紅」、「達麗台中高鐵案未公開就賣 6 成 董座驚：銷售速度比高鐵還要快」、「下個護房神山是它？李同榮：房市將『見軌則發』」、「高雄首件公辦都更案！「鳳山都更好」單坪挑戰 4 字頭」、「大開發案進場 台中商辦總銷逾千億」、「台中七期商辦 日大廠進駐」、「抓住趨勢 鄉林打造全齡宅」、「熱點出擊 京城蓋頂級 A 辦」、「青年房貸族 每人平均背 862 萬元」、「新屋齊發 全台推案 4,552 億」，其中僅一篇為負面報導，其餘報導皆屬正面消息，因此讓該日情緒分數較高！

【聯合報】每日情緒彙總



● 每日正面情緒

先將資料欄位進行篩選，透過「`sentiment_value > 0`」條件式，篩選出情緒為正的資料，再以「日」為單位，將該日情緒彙整分數進行加總。在近一年半的期間中，有 263 天出現正面情緒報導，正面情緒分數如下圖所示。

分群彙總 (每日正向情緒) (166)

參數設定

Input - 182

任務結果

選擇日期欄位 *

artDate

選擇日期類型 *

日

日期格式 *

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

匯總函數 *

sum

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

negative_count

negative_words

booster_count

booster_words

negator_count

negator_words

sentiment_value

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

system_id

Unnamed: 0

artTitle

artUrl

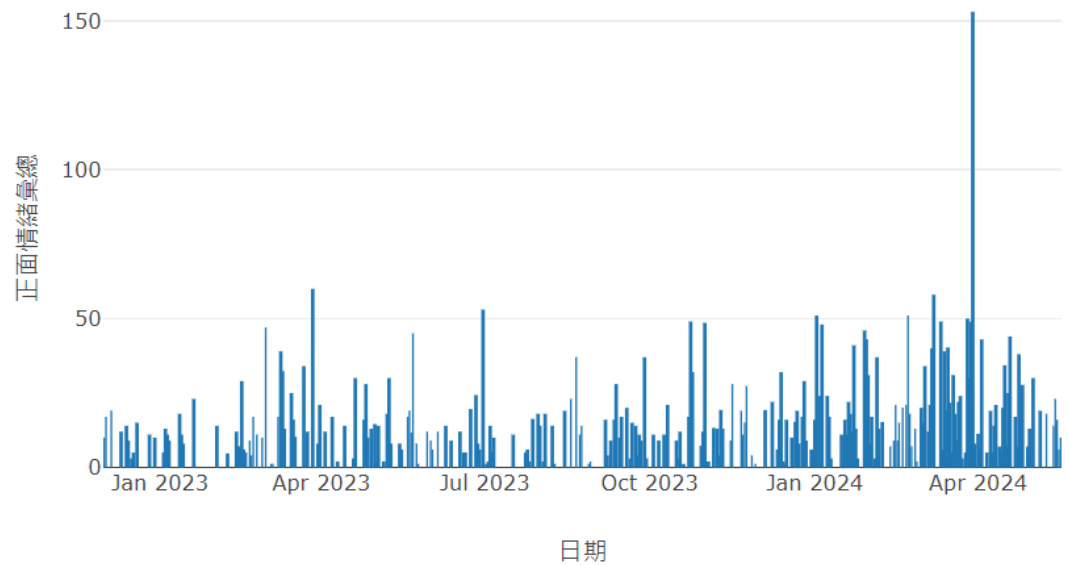
artDate

artCategory

artContent

儲存更改

【聯合報】每日正面情緒彙總



● 每日負面情緒

先將資料欄位進行篩選，透過「`sentiment_value < 0`」條件式，篩選出情緒為負的資料，再以「日」為單位，將該日情緒彙整分數進行加總。在近一年半的期間中，有 59 天出現負面情緒報導，副面情緒分數如下圖所示。

分群彙總 (每日負面情緒) (167)

參數設定 Input - 183 任務結果

選擇日期欄位 *

artDate

選擇日期類型 *

日

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

negative_count

negative_words

booster_count

booster_words

negator_count

negator_words

sentiment_value

日期格式 *

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

匯總函數 *

sum

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

system_id

Unnamed: 0

artTitle

artUrl

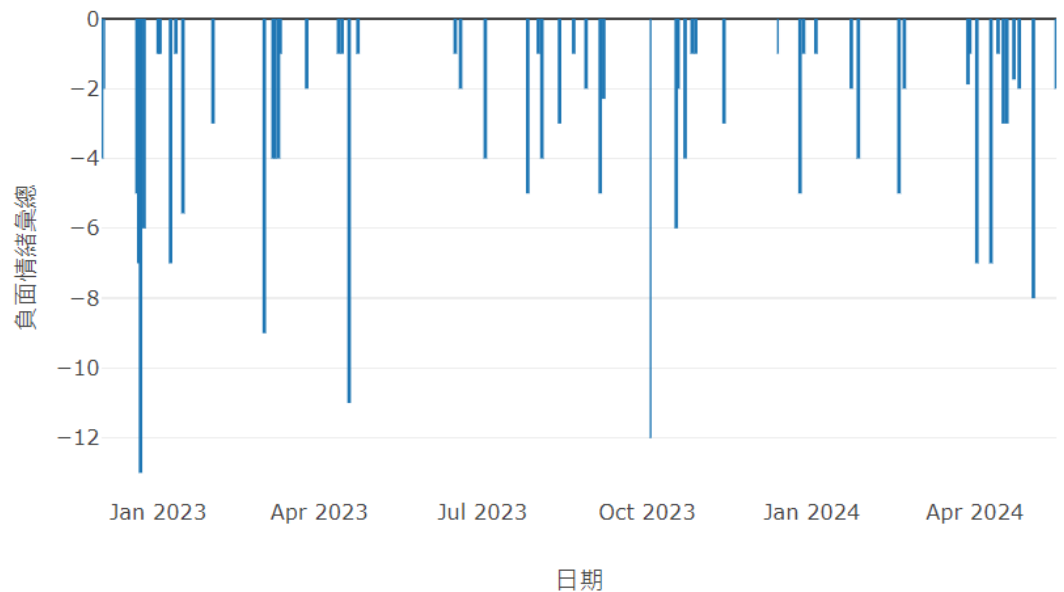
artDate

artCategory

artContent

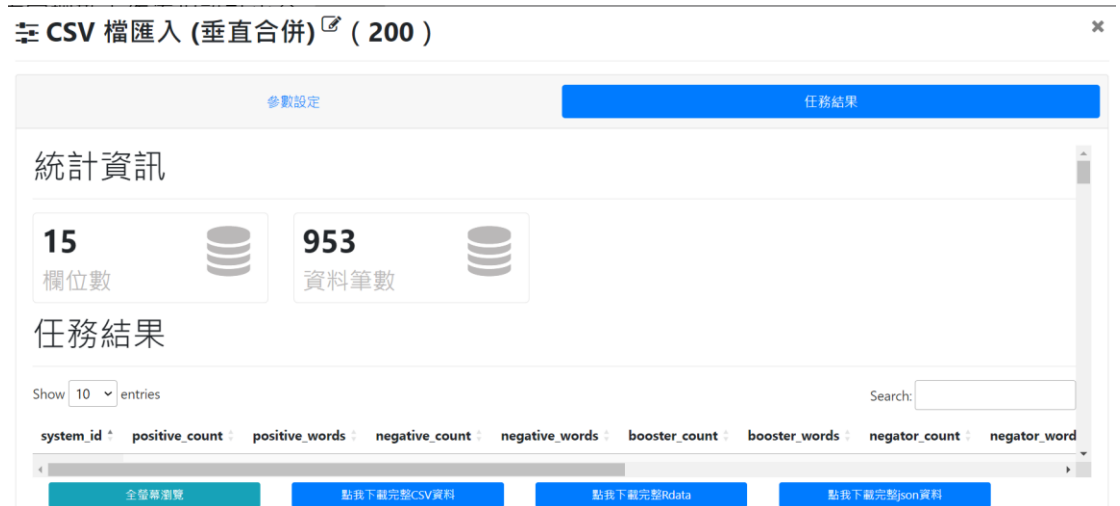
儲存更改

【聯合報】每日負面情緒彙總



6. 垂直合併

將自由時報、東森新聞及聯合報的資料整併後匯入，共有 953 筆資料。



7. 資料預處理(替換字串、中文斷詞、清除停用字等)

(1) 替換字串

≡ 替換字串 (205)

參數設定

Input - 200

任務結果

選擇處理欄位 *

artContent

替換字串設定 ⓘ

高雄>> 高雄市

選擇替換規則檔案 ⓘ

-----請選擇-----

儲存更改

(2) 中文斷詞

中文斷詞延續自情緒分析時所使用的斷詞結果，加入科技之心、高雄市長、馬稠後產業園區等詞彙。

≡ 中文斷詞 (206)

參數設定

Input - 205

任務結果

選擇處理欄位 *

result

定義詞彙 ⓘ

科技之心 100
高雄市長 100
馬稠後產業園區 100
1.4奈米 100
2奈米 100

選取字典 ⓘ

-----請選擇-----

儲存更改

(3) 清除停用字

並透過詞頻計算及網路圖的呈現結果清除停用字，如今年、表示、目前、未來、去年、近年、指出、成為、包括、位於、認為、預計等。

≡ 清除停用詞 (207)

參數設定

Input - 206

任務結果

清除英文字母 *

否

清除數字 *

是

清除換行符號 *

是

清除特殊標點符號 *

是

清除html tag *

是

自定義停止詞

今年
表示
目前
未來
去年

儲存更改

8. 轉換為 DTM

將資料轉為 DTM 格式，篩選詞彙數量設定為 200。

轉為DTM (204)

參數設定

Input - 207

任務結果

保留詞彙
以換行符號區隔，e.g.
國立中山大學
西子灣
壽山...

最多篩選詞彙數量
200

儲存更改

9. 視覺化呈現

(1) 關聯式文字雲

使用預設參數開始執行，從關聯字文字雲中可以發現，大部分字詞被分為兩群，分別為上方與房市及房價相關的文字(淺藍色文字)，以及下方對於住宅、政府、生活圈等都市發展相關字詞(深藍色文字)。

關聯式文字雲 (shiny) (226)

參數設定

Input - 204

任務結果

分群數
20

迭代次數(最少250次)
1000

聚合演算法
single

文字雲顯示字數
1000

距離計算公式
euclidean

儲存更改

關鍵字網路分析工具

詳細設定

選擇詞類

others

選擇次層詞類

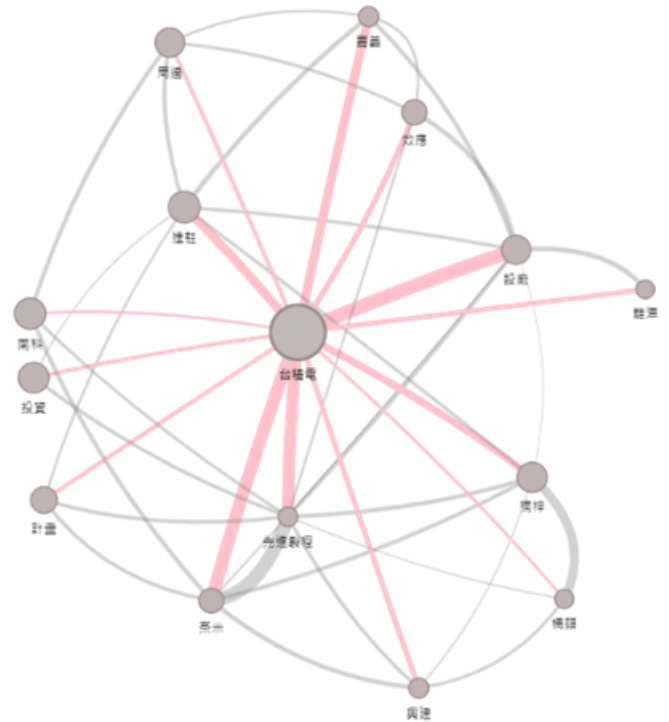
台積電

設定網路節點數量

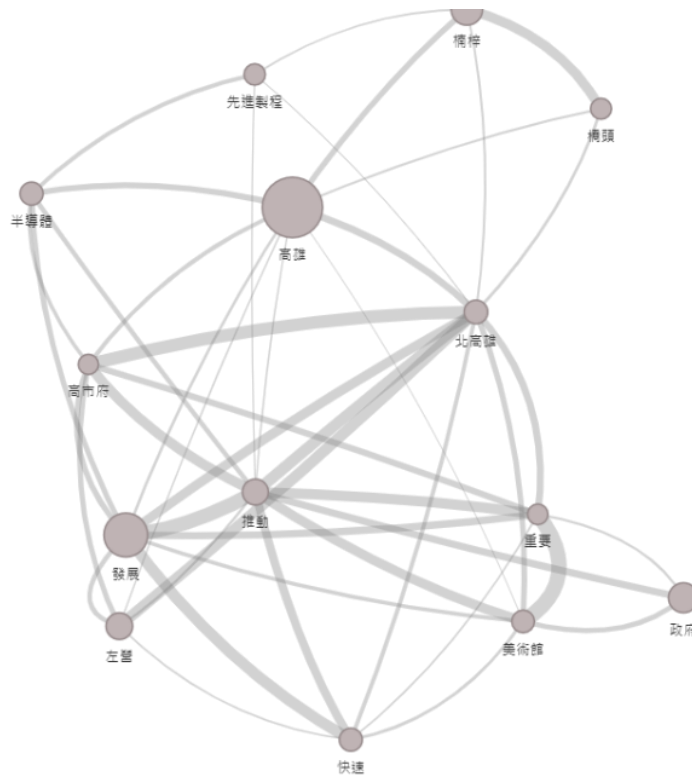
41532

依關聯強度篩選鏈結

00.51

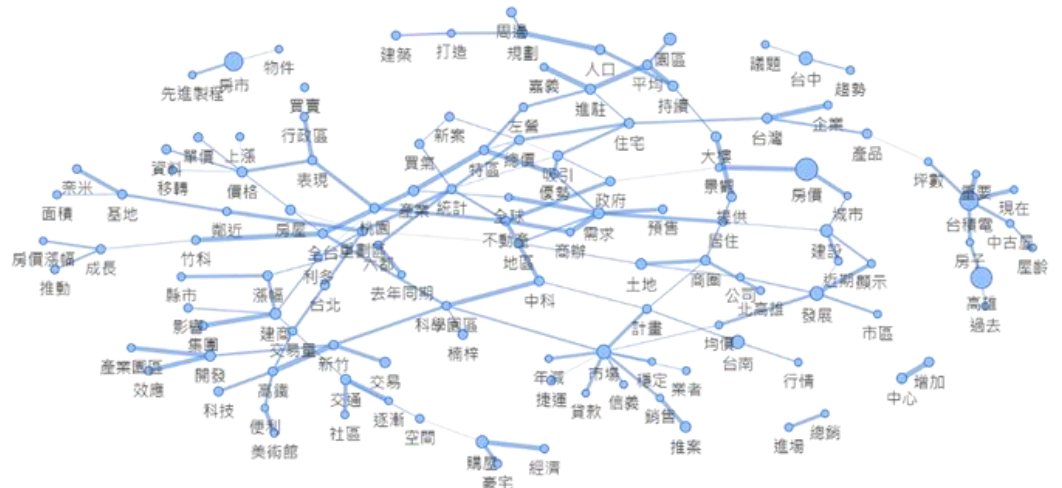
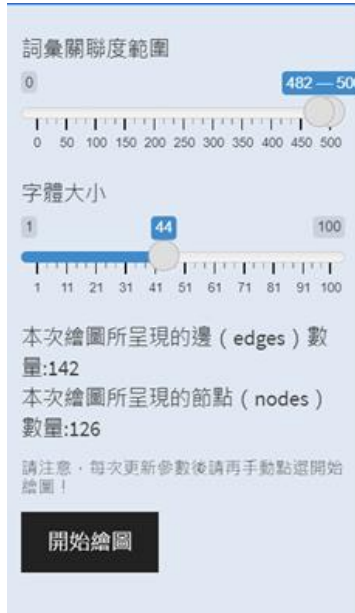


將次層詞類設定改為高雄，其他參數不變下，發現其關聯字包含地區名稱，包括楠梓、橋頭、左營及美術館，並且也與半導體高度相關。



(3) 單中心網路圖

將詞彙關聯度設定在 482-500 之間，edges 數量 142，節點數量 126。



F. 結論

本研究以自由時報、東森新聞與聯合報等台灣三個重要新聞媒體，在2022/12/01~2024/05/19 近一年半之間，透過蒐集與台積電相關的房市新聞，進行文字分析，了解在相同時間區間中，是否會因不同媒體而有不同情緒之報導，此外，也愈透過視覺化方式，了解字詞之間的關係。

在情緒分析而言，若進行每日情緒彙整，可以發現自由時報報導最為正面，有 95.36% 的日期為正面情緒報導；其次為聯合報，有 94.22% 的日期為正面情緒報導；最後為東森新聞，與前兩者差異較大，僅 82.8% 的日期為正面報導。若從上述結果來看，似乎自由時報與聯合報兩者間差異不大，皆有 95% 上下的日期為正面報導。

然而若改以每日負面情緒進行彙整，雖然順序上一樣沒有改變(自由時報最少，其次為聯合報，最後為東森新聞)，但占比則有較大的差異：自由時報僅有 9.3%的日期有負面情緒報導，聯合報則有 20.1%的日期有負面情緒報導，東森新聞則有 26.6%的日期有負面情緒報導，此時聯合報會較接近東森新聞。透過觀看資料後可以發現，聯合報共有 25 天同時出現正面與負面情緒

之文章，且當日正面情緒之分數皆大於負面情緒，才会有此現象發生。

在資料圖像化的部分，我們可以透過關聯式文字雲及單中心網路圖、字詞網路圖分析，進一步探討媒體對於台積電前往高雄設廠並進一步帶動都市發展的論述。從關聯性文字雲中，我們可以發現高雄、台積電及房價這三個字詞是文章加總後出現頻率最高的，而圖中與高雄鄰近的字詞包括楠梓、橋頭等地點，也就是台積電宣布設廠之後房價逐漸攀高的地區，其他圍繞在高雄周遭的字詞則提及產業、發展、需求、建設、交通、政府及住宅等，推估為媒體對於台積電設廠高雄後對於都市發展的相關議題討論。另一部分的字詞則以房價為核心，相關字詞包括購屋、市場、交易等，特別的是這一群文字雲中出現了數個地區，包括南科、竹科、中科等科技園區，應是作為與台積電進駐高雄後的發展模板及比較案例。

在單中心網路圖的部分，我們以台積電為核心字詞向外延伸，可以發現先進製程、奈米、設廠、進駐等詞彙，以及地點如楠梓、橋頭、南科、龍潭、嘉義等，由高雄為核心字詞向外延伸，則得到高雄市內的各個地區名稱以及半導體的關鍵字，得以解釋台積電、房價與高雄三者之間的脈絡。

從字詞網路圖觀察字詞之間的關係，可以在圖中最右側發現台積電及高雄，而延伸出去的高關聯性字詞則圍繞在房子、中古屋、房價及坪數等。

綜整上述而言，於相同資料期間，房地產新聞雖然多數仍為正面情緒報導，然而在實際上，不同新聞媒體間則會有類似民意調查時所產生的「機構效應」，因選取資料期間中央政府係由民進黨進行執政，政治光譜上偏綠的自由時報會有較多的正面報導，而政治光譜偏藍的東森新聞與聯合報，則會有較多負面新聞產生，若民眾在觀看新聞時選擇不同媒體，可能就會接受到不同情緒，進而影響其大選時政黨與候選人之偏好。

而從關聯字文字雲、單中心網路圖、字詞網路圖等資料視覺化方式，可以發現在近一年半以來，台積電對高雄房市有相當大的影響，文章中常與楠梓、橋頭等高雄相關地點同時出現，若鎖定高雄作為單中心網路，則會出現產業、發展、需求、建設、交通、政府及住宅等，顯現台積電於高雄楠梓設廠消息傳開後，明顯帶動附近房市發展與需求。

G. 研究限制與建議

1. 本研究係以「日期」為資料分割方式，進行情緒分析，若欲更加細緻了解三大媒體是否會針對相同主題會有不同報導差異，則須以「主題」來進行資料分割，因為相同主題的內容可能會分散在不同日期間，例如台積電於楠梓設廠，文章涵蓋日期就可能會超過一年。
2. 不同新聞媒體資料筆數差異較大，其中又以東森新聞資料筆數較少，僅有 90 筆資料，且於 2023 年 6 月起即無資料，此外聯合報房地產版也無法於 Tarflow 上直接取得資料。經了解各大新聞媒體網站後猜測，可能因新聞網站有更動、房地產主題會有另設網頁(並非與其他版)等因素，Tarflow 爬蟲上才沒有資料。而本次研究因受時間限制，發現此問題較晚，已經來不及補充東森新聞近一年的資料，在資料比較與分析上可能會有所偏誤。後續若還有時間，應補充相關資料，才能完善研究內容。

H. 參考資料

1. [「護國神山」成房價保證，台南 2 區漲幅更勝新竹；南科周邊產值、房市雙漲](#)
2. [台積電設廠帶動房價飆升！一張圖看先進製程布局帶動房價漲幅](#)

I. YOUTUBE 連結

<https://youtu.be/KUBwPYam3RY>