# 社群媒體分析 期末報告

一、 題目:大選後房地產輿情文字分析

二、組別:第八組

三、成員:

組長: N114320016 李儼哲

組員:N114320007 黃顗哲

組員:N114320010 黃善羚

組員:N114320031 洪宜綾

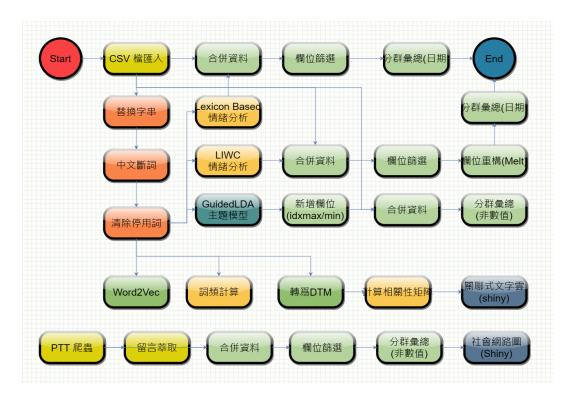
# 四、前言

今年總統大選後幾家歡樂幾家愁,每位政黨的候選人在 選舉期間都提出不少與房市相關的政見,例如:

- 柯文哲是大力蓋社宅
- 賴清德提出社會住宅+包租代管 50 萬戶 & 333 優惠
- 侯友宜的政策則是公宅多元六政策

我們這組想研究大選過後有關於房地產輿情相關的情緒變化,並使用情緒分析、主題模型、社會網路分析之。

### 五、 基本資料



上圖為我們 Tarflow 的流程圖,資料是從 DCard 房屋版 抓取下來的,時間設定 2024 年 1-3 月,經過分群彙總可以 得知一月文章有 804 篇、二月文章有 758 篇、三月文章有 985 篇。



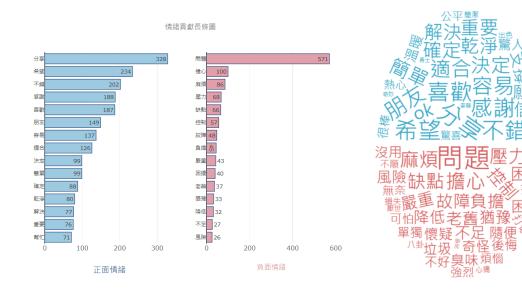
# 六、詞頻計算

我們可以看到第一名的關鍵字是台中,可見台中的房市 討論度在全國裡是最高的,其他還有建商、重劃區、空間、 車位、公園、捷運、陽台、大樓,這些在找房子時會被考慮 進去的點,其中因為六都的房價高規劃的重劃區多建案多, 導致重劃區也是熱門的關鍵字,裡面最讓人意外的就是床墊, 看來現代人很重視生活品質,導致床墊也變成熱門關鍵字。

# 七、Lexicon Based 情緒分析

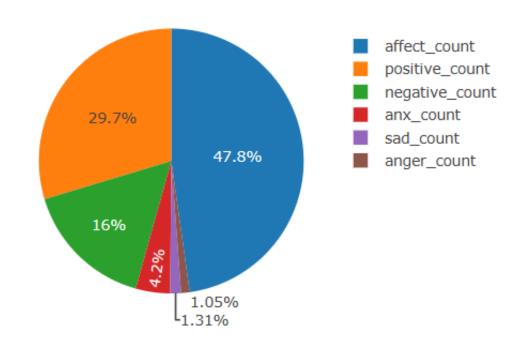
正面的情緒字有喜歡、適合、決定、乾淨、簡單、溫暖、 朋友,反應了現代人在選擇住家時考慮的點偏向簡單不喜歡 複雜的環境。

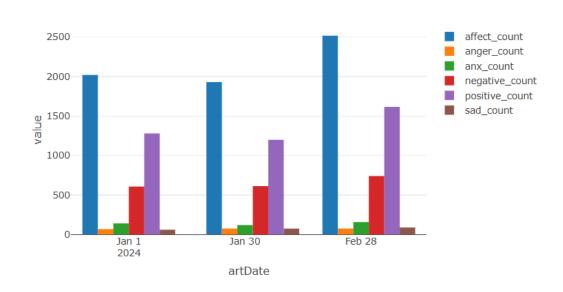
負面的情緒字可以發現問題特別的高、甚至比正面情緒 的第一名還要高,這強烈顯示在居住這塊對於各種問題特別 害怕與關注的,其他還有擔心、麻煩、缺點、故障、老舊。



# 八、LIWC 情緒分析

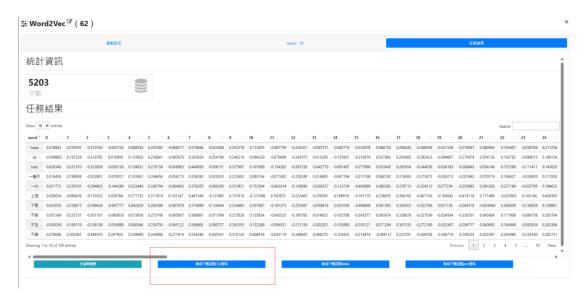
我們可以從圓餅圖看出 affect 占 47%、anger 占 30%、anx 占 16%,而各情緒的比例變化不大;從直方圖看出三月份因為文章較多,導致各情緒高度也比較高。





### 九、Word2Vec

我們使用 mobile01 的居家房產來做這次的 Word2Vec 設定參數為 100 維後,轉為關鍵字詞 word 跟每一個詞之間的相似度並點選下載完整 csv 資料。

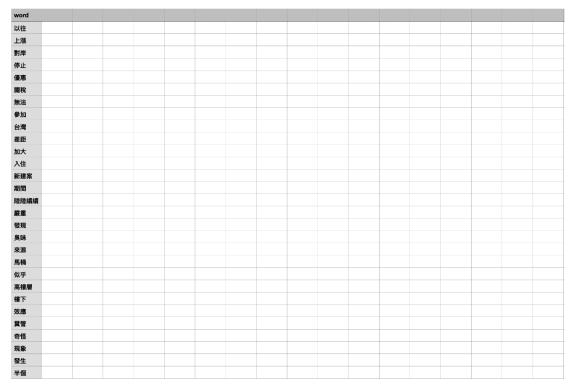


下載後的 csv 資料轉分別轉為兩份 TSV 檔分別為 Embedding 及 metadata。

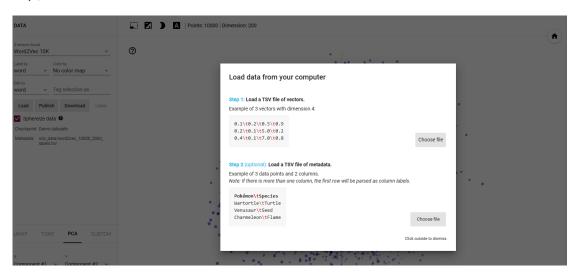
# Embedding

0.011702917	0.17013207	-0.26703745	0.052268654	0.14088264	0.2421147	-0.045336016	0.34383938	-0.016927002	0.15165403	-0.0029418168	-0.034034047
0.026054207	0.06643945	-0.15335211	0.026783977	0.27173194	0.31191888	0.10314731	0.48154795	-0.12768106	0.1576104	-0.1215084	0.19292517
0.021101153	0.20861144	-0.299873	-0.018323896	0.106977016	0.25272572	-0.054274436	0.3756537	0.02959045	0.24040668	0.10770097	-0.08527508
0.040461387	0.22548693	-0.30793384	-0.11048803	0.046928402	0.2620365	-0.02967611	0.36088324	0.044375006	0.2787741	0.16171046	-0.10964877
0.036558244	0.22384515	-0.27445105	0.091900975	0.20612198	0.2061548	-0.02996581	0.39693826	-0.07699221	0.1660053	0.0024123476	-0.0043440047
0.020134069	0.17337494	-0.26131654	-0.007816905	0.117783345	0.25227782	-0.026286585	0.36520088	-0.001439159	0.20239766	0.04957583	-0.030426275
0.016938152	0.049931876	-0.22225969	-0.13918553	-0.020777412	0.3264098	0.048958417	0.2382501	0.07393554	0.21933456	0.20302533	-0.11019576
-0.0016179117	0.2011631	-0.2860119	0.027002417	0.13783099	0.23289655	-0.04993394	0.37282038	-0.009802765	0.20489433	0.02397145	-0.066261806
0.245728	0.22687022	0.043362603	-0.08361345	0.1498682	0.4265638	0.14996703	0.4650562	-0.12751652	0.22255115	-0.073169574	0.30564806
0.0011008453	0.16652413	-0.23834524	0.009016183	0.17383172	0.25531456	0.0029007632	0.39867795	-0.021721669	0.19935581	0.02767989	-0.055658877
0.04039899	0.23414806	-0.33611208	-0.017021444	0.101043455	0.23043482	-0.06705639	0.40154803	0.033898856	0.26980108	0.09706555	-0.058888886
-0.047889315	0.12090685	-0.26722124	0.06431063	0.192646	0.25743097	-0.053330433	0.38889948	0.018838897	0.17712961	0.018408012	-0.063097574
-0.10038314	0.09367082	-0.30610657	0.2781921	0.39223495	0.29714105	-0.063451916	0.4885944	-0.009609541	0.1142541	-0.11537376	-0.015161488
0.00028401808	0.14217757	-0.2571976	-0.011426617	0.12595043	0.24471857	-0.022914369	0.34022406	-0.00543823	0.18510087	0.05269779	-0.057462
-0.007899427	0.1564749	-0.28403753	0.047487006	0.16135526	0.2396032	-0.04637019	0.36958104	-0.0008688259	0.18556902	0.026260138	-0.048674587
0.018371701	0.13794078	-0.31332394	-0.0911621	0.034732115	0.2806852	-0.039788593	0.32226658	0.07212199	0.26375815	0.16770798	-0.091935255
0.04897639	0.15296185	-0.44415906	0.0422372	0.1371442	0.38143188	0.013274904	0.37790585	0.1188109	0.28310043	0.33961138	-0.20112295
0.022488805	0.20669042	-0.33251047	-0.051405374	0.073971294	0.2676138	-0.057622805	0.36330047	0.064868614	0.26838437	0.15301146	-0.12351321
0.015878383	0.17133553	-0.28410047	0.02216325	0.13517909	0.25599143	-0.028210064	0.36398938	0.00035826885	0.1946273	0.042246535	-0.057322383
0.02785884	0.25622666	-0.483034	-0.15871316	0.02461954	0.33169016	-0.027375497	0.44544068	0.19115311	0.44588277	0.34636974	-0.28152308
-0.0074341875	0.20328441	-0.32734945	0.12852554	0.20416036	0.29955184	-0.089170195	0.39449564	0.06051111	0.19313426	0.069113635	-0.09376555
-0.11761946	0.058063462	-0.2895864	0.16991402	0.33959693	0.24463011	-0.006117521	0.50109327	-0.020649087	0.18307829	-0.095812805	-0.03709506
-0.049778335	0.09465243	-0.3268147	-0.025336798	0.07427448	0.3302553	-0.014344576	0.30363157	0.13424511	0.2151068	0.17725566	-0.19935384
-0.006856816	0.07628855	-0.23242225	0.07966251	0.22524199	0.23841585	0.02286364	0.38436967	-0.08169833	0.13358794	-0.09467014	0.04976693
0.018494563	0.16610387	-0.30219603	-0.031191837	0.104411244	0.2508613	-0.03322152	0.35288805	0.03277774	0.22105156	0.0965334	-0.08467933
0.023987802	0.2109535	-0.3477728	-0.05368955	0.090910316	0.26777545	-0.06428872	0.39014134	0.07150382	0.2860523	0.17109883	-0.11526742
-0.0053281225	0.105470024	-0.23531914	0.016688775	0.16154091	0.2606634	0.011486676	0.3498354	-0.023777837	0.17715381	0.02593458	-0.0402299
0.01596366	0.053445455	-0.26025334	-0.08863833	0.0710098	0.34068894	0.036425784	0.3110048	0.059898056	0.24177913	0.15932916	-0.060859226
-0.0024161271	0.12517431	-0.25954905	0.0026235392	0.13747352	0.24736203	-0.02124105	0.3593417	-0.009964139	0.193084	0.033190873	-0.03692222
0.021772642	0.25978106	-0.29490203	-0.14429557	0.023439694	0.28070432	-0.08460313	0.37826544	0.060294993	0.3519206	0.1525643	-0.082314074

# metadata



再到 Embedding projector - visualization of high-dimensional data (tensorflow.org), 將剛剛兩個檔案分別匯入。

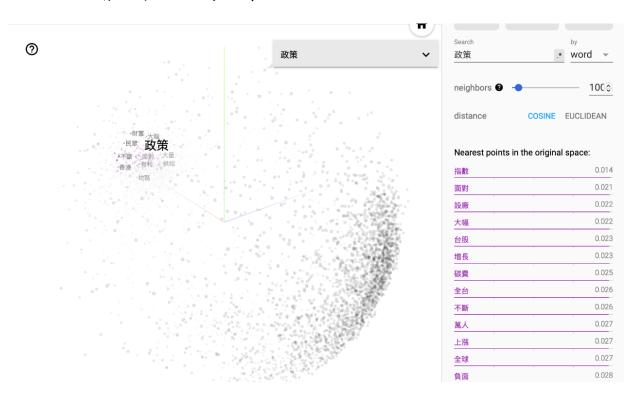


在這次資料蒐集中發現經濟、大選牽動房市!2024年房市熱門關鍵字揭曉在最熱門關鍵字上,包括「炒房」、「碳權」、「買氣」都是網友們最在意的關鍵字,除此之外「政策」、「央行」、「升息」,亦吸引了不少網友矚目。其中,「政策」、「炒

房」兩大關鍵字有其關連性,在「打炒房」主軸下,跟以往似乎有不同的詮釋,之前主要討論打房政策影響,以至於「地權」內容特別突出,檢視其中討論。多為2023年7月上路的「平均地權條例」相關問題,顯見以往民眾還是認為政策有助於房價修正。到了2024年,在政策作多下,交易量谷底反彈,包括買氣、炒作、央行等字眼浮上檯面,讓市場呈現不一樣的風貌。

所以在本次分析中,我們從房屋買賣文章中提取了與 "政策" "炒房" "買氣"相關的關鍵詞,並使用主成分 分析 (PCA) 技術來探討它們之間的關聯性。

# • 將關鍵字設定為政策



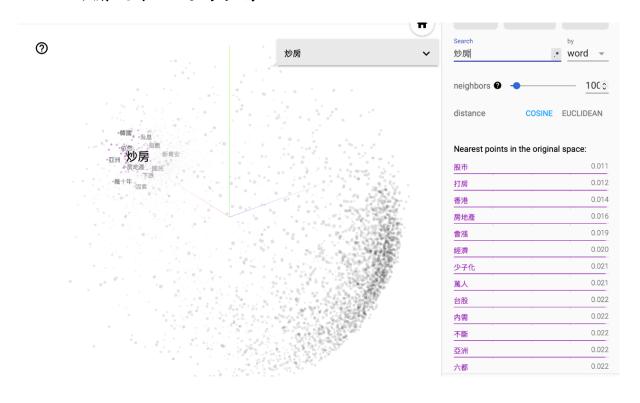
可以發現在此期間,政府推動了一系列房地產調控措施, 房地產市場的討論聚焦於一系列關鍵政策和經濟指標,這些 因素交織在一起,影響著市場的未來走,以穩定房價和促進 經濟增長。這些措施包括調整稅收、放寬房貸條件及增加土 地供應。隨著政策實施,台股指數波動反映了市場對政策的 反應,成為房市走勢的重要指標。

政府同時推動設廠計劃,吸引國內外企業投資,帶動周邊房地產需求大幅增長,工業區房價上漲成為市場熱點。然而,全球經濟不確定性和貿易環境變化使投資者更為謹慎,房市增長有所放緩但依然穩定。

此外,碳費政策推動了綠色建築的發展,吸引大量買家 關注環保和能源效率的新項目。房地產專家顏炳立指出,這 些政策有利於市場健康發展,但投資者需警惕風險。

儘管面臨挑戰,台灣房市仍吸引大量投資者,許多人希望在新政策的利好下獲得理想回報。然而,房價快速上漲和市場過熱帶來了一些負面影響。

# • 將關鍵字設定為炒房



2024年的台灣總統大選正如火如荼地進行中,房地產市場成為熱議話題。關於"炒房"的討論尤為激烈,反映出市場對政策和經濟走向的高度關注。

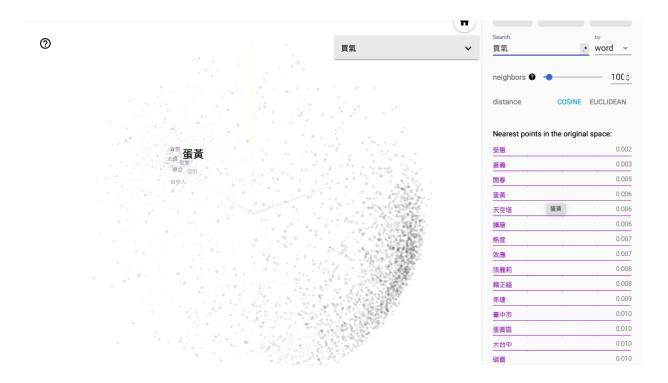
隨著大選臨近,**股市和台股**的波動成為市場情緒的晴雨表。投資者密切關注政府的**打房**政策,擔心這些措施會影響房地產市場的熱度。然而,許多人認為,即使政府採取強力的打房措施,房價依然會上漲,因為內需市場強勁。

在這場大選中,台灣的房地產市場與亞洲其他地區,如 香港,也被拿來比較。香港的房地產市場長期以來受全球經 濟影響,而台灣的市場則面臨不同的挑戰,包括少子化和內 需變化。這些因素共同作用,影響著市場的供需平衡。

同時,台灣的**六都**(台北、新北、台中、台南、高雄、桃園)成為投資者關注的焦點。這些城市的房價增長預期受到多方因素影響,包括經濟發展、人口流動和政府政策。儘管面對挑戰,投資者對六都的房地產市場依然充滿信心,認為這些城市的房產具有長期的增值潛力。

在全球經濟不確定性的背景下,亞洲市場尤其是台灣的 房地產市場表現出了相對的穩定性。這種穩定性部分歸因於 台灣的經濟政策和內需市場的支持。儘管面對全球經濟挑戰, 台灣的房地產市場不斷適應變化,展現出強勁的增長潛力。

### • 將關鍵字設定為買氣



2024年台灣總統大選期間,房地產市場的"買氣"成為 熱議話題。隨著選舉的臨近,市場的熱度和購房需求顯著上 升,許多地區的房市表現引人注目。

首先,大選期間的政策利好帶來了市場的**受惠**效應,許 多投資者和購房者對房地產市場充滿信心。嘉義這樣的城市 在大選前夕表現出色,成為房市的新熱點。隨著**開春**來臨, 市場的活力進一步增強。

台中市的**蛋黄區**和**大台中**地區,憑藉其優越的地理位置和發展潛力,一直是投資者的焦點。另外是位於台北市中心的**天空塔**建案,吸引了大量的目光,成為市場的明星項目。隨著更多企業在台中**擴廠**,當地房地產市場的**熱度**不斷攀升。

著名房地產專家**賴正鎰**和**張蘿莉**在各大媒體上頻頻露面,他們分析了市場的動態並預測未來的發展走向。賴正鎰 指出,大選帶來的政策變動和經濟發展對房市有積極影響, 而張蘿莉則強調了市場的內在潛力,特別是大台中地區的發 展機遇。

此外,市場中還出現了"**串聯**效應",即不同地區的房市表現相互影響,共同推動了整體市場的繁榮。台中市作為這種串聯效應的核心,帶動了周邊地區的房市活躍。

政府推行的**碳費**政策也成為討論的焦點。雖然這一政策 旨在促進環保和可持續發展,但對於房地產市場而言,卻帶 來了新的挑戰和機遇。開發商們開始更加重視綠色建築和節 能措施,這些新趨勢進一步影響了購房者的需求和市場的走 向。

# 十、 關聯式文字雲 (shiny)



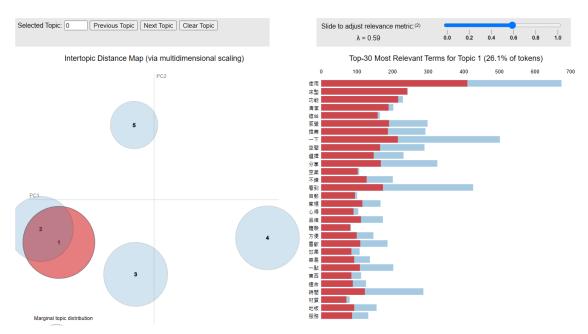
我們使用關聯式文字雲分成五群,可以發現左邊灰色主要是在講有關於台中的房市,右邊紫色主要在講的就是住家空間配置、修繕、家具相關的字詞。

# 十一、GuidedLDA 主題模型

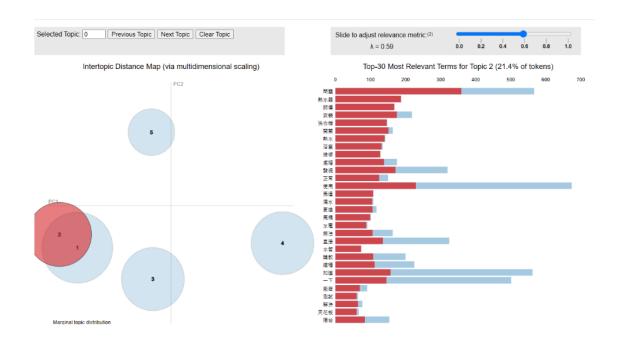
我們根據字詞內容先抓出並設定台中與高雄兩個主題 種子字,並把主題數設為5,以此觀察主題分類的效果。



第一個關鍵字有床墊、清潔、空間、品牌···等,主要關鍵字集中在家具的討論,因此我們定義為群居家空間群。

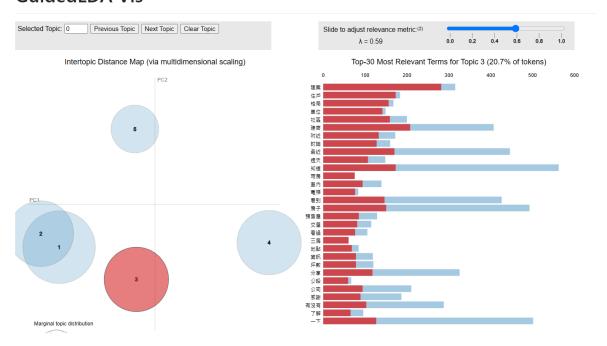


第二個群與第一個群分常接近,代表他們之間的關聯度 非常高,關鍵字有問題、師傅、維修、處理、更換…等,主 要在討論家中設備更換維修,故我們定義他為居家修繕群。



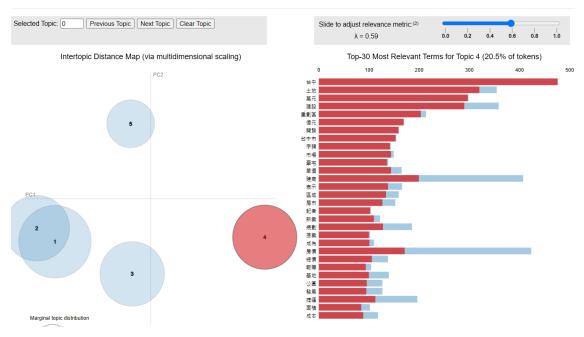
第三個群相對於第四第五離第一第二群比較接近,我們 也可以從相關詞彙看出,建案、格局、社區、建商、透天… 等,我們可以看出這個群是關於找房、看房、建案的討論, 固定義為看房群。

#### **GuidedLDA Vis**



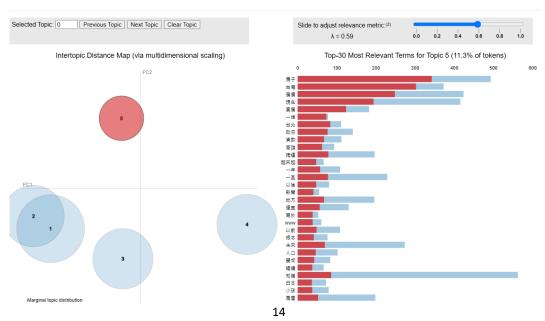
第四個群就是我們種子字裡定義的台中群,這邊主要都 是台中房市的討論,包括了重劃區、開發、豪宅…等。

#### **GuidedLDA Vis**

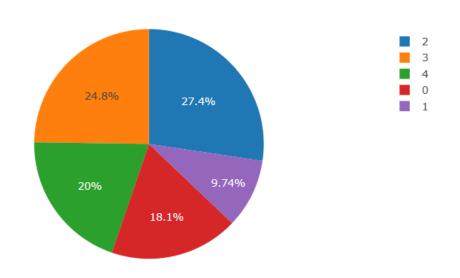


第五個群就是我們種子字裡定義的高雄群,主要是在討論高雄房價,關鍵字有貸款、新聞、現在、買房、政府,可以從討論得知隨著台灣的政策讓台積到了高雄,使著高雄房價節節升高,討論熱度也提升,而"現在"買房不然房價會一直上升相關議題也不斷發酵。

#### **GuidedLDA Vis**

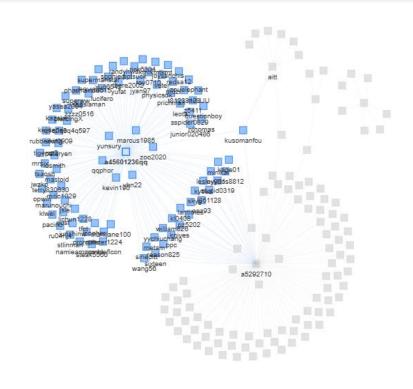


我們也對分出來的五個群進行分群統計成圓餅圖,可以 發現群居家空間、居家修繕群、看房群、台中群、高雄群的 占比分別為 27%、24%、20%、18%、9%。

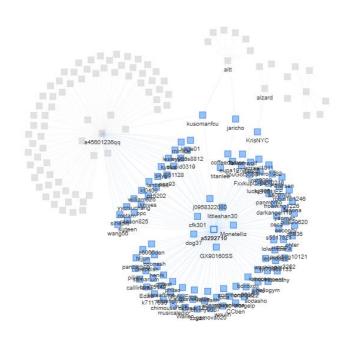


# 十二、社會網路圖 (Shiny)

透過社會網路圖可以看出節點之間的互動和連接方式更加明確,整個網絡可以分成四個主要群體,其中三個群體是發言者,一個群體是回文者。以下是對這些群體的簡要說明:



發文者指出華固建設(2548)在股東會上,現階段房市景 氣熱絡且供需平衡,但需警惕未來的反轉可能性。只要不過 度推案,房市景氣可維持 1-2 年。華固建設通過了每股配發 現金股利 7.5 元、股票股利 1 元,並計劃 2024 年在台北及 台中推出多個大型建案,合計達 470 億元。他提到,台灣資 金充裕,國內經濟高成長、股市創新高,市場交投熱絡,各 大產業向上成長將帶動房市發展。房地產市場目前處於繁榮 階段,但需避免超額供給。預計推出的 6 個新案總銷超過 450 億元。發文者認為,受平均地權條例、房地合一及土建融政 策影響,大建商調控推案數量,變得更保守,新案一出即被 搶購,購房者需更加積極。發文者認為,由於平均地權條例、 房地合一以及土建融政策的影響,大建商在調控推案數量, 變得更保守。因此,市場上新案一出就被搶購,導致想買房 的人需要更加積極。 回文者討論內容涉及對不推案和現金分配的看法,部分股東希望在不推案時多發現金股利,認為政府的政策如土建融限貸阻止了大量推案。政策影響方面,禁止預售屋轉售和實登政策被認為是阻止建商聯合抬價的措施,打炒房政策如禁紅單和房市信用管制也引發爭議。房價與股市的關聯被提及,認為股價下跌會影響房價停滯或小跌。紅單與預售屋轉讓問題強調紅單應該禁止,但預售屋合約轉讓是另一回事。打擊小建商的政策被認為旨在減少推案量、降低價格競爭以穩定房價,但也有觀點認為這些政策過於嚴苛,對市場有負面影響。



主要討論日前輝達執行長黃仁勳表示,台北、台南、高雄是投資研發中心的選項,並指出桃園市為吸引輝達 (NVIDIA)投資研發中心的選址所做的努力。發文者認為,無論輝達最終選址何處,只要設在南港或新北板南線上,對青埔房市都是利好。

針對輝達(NVIDIA)投資研發中心的選址議題,留言者 主要表達了以下觀點和意見:

### 1. 地點選擇的討論:

- 一些人提到桃園的不同區域,如龜山、龍潭、楊梅, 並討論這些地區是否適合輝達的投資。
- 有人認為桃園不太可能成為輝達的首選,並建議輝達 選擇北士科(新北市新店、土城、樹林、三峽地區) 或新竹。

### 2. 對桃園市長張善政的看法:

- 有人批評張善政,認為他對龜山區的態度消極,並對 他的治理表示不滿。
- 有人諷刺張善政和市府的努力,認為他們爭取輝達投資是不切實際的。

# 3. 其他縣市的競爭:

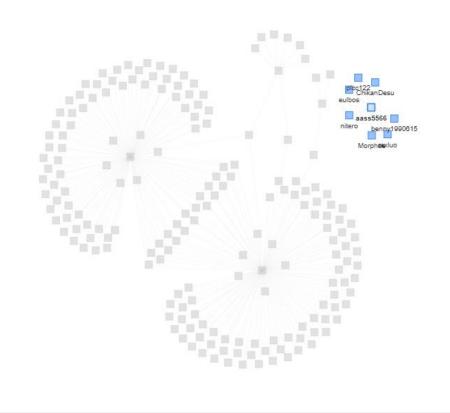
- 討論中提到除了桃園外,台中等其他縣市也在爭取輝 達的投資。
- 有人表示台中在這方面的準備不足,甚至有嘲諷台中 主要從事詐騙活動的言論。

# 4. 對輝達的建議:

- 有人建議輝達在台灣各地廣設研發中心,而不僅僅集中在一個地區。
- 有人提到輝達應該選擇有高科技人才的地區,如新竹, 因為當地有清華大學和交通大學。

### 5. 對桃園市未來發展的看法:

- 有人認為桃園市的地理位置不錯,但目前的發展和資源不足,不適合吸引輝達這樣的大企業。
- 一些留言者對桃園市府的努力表示不屑,認為他們的 爭取行動只是蹭熱度,並不會成功。



發文者對居住正義的議題表示擔憂,認為台灣房價不斷 上漲,將來可能會達到 50 萬一坪,類似於香港的高房價。這 種情況下,一房成為主流,許多人窮極一生也買不起房。發 文者認為,這樣的高房價會迫使人們到國外買房置產,以實 現居住正義。儘管這樣可能會被批評,但如果能在國外住得 像個人,這也是一種解決方案。

針對居住正義和到國外買房創業的議題,留言者表達了 多種看法。一些人提到某網紅在泰國買房後,外國人社區單 價上升,並表示許多人學泰語移民泰國視其為離開台灣的選擇,但也有人認為泰國高薪工作難找,生活品質不如台灣。在創業和生活質量方面,有人鼓勵在泰國創業發大財,但也有人認為在台灣無法買房的人到更先進國家生活會更困難,重點應是賺更多錢。對於持有成本和房地產,有人認為台灣住宅用房地產因持有成本低是優良資產,另有人擔心去國外買房也會成為炒房者。最後,對於不同國家的創業比較,有人質疑在泰國創業的必要性,認為創業的條件和挑戰在各國相似,出國創業的理由不明確,創業本質無太大差異。

留言者對於以下內容做出許多回覆像是高雄小港透天便宜的原因表示,因為小港和高雄市區被機場切開來,發展困難。對於鳳山地段,回文者指出該地區有衛武營、百貨公司、交流道和捷運,並討論了鳳山的都市更新和建案名稱。針對八卦版政治化的問題,回文者認為,不只是綠的爛,白的也沒多好,主要談論房產到八卦的涉略。

# 十三、 Rag

Step 1:於 Tarflow 上下載建案檔案(.csv)

Step 2: 將(.csv)轉換成(.txt)

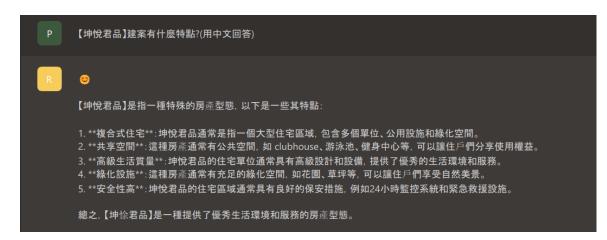
Step 3: 將(. txt)透過相關網站轉換成(. pdf)

Step 4: 將(.pdf)檔案上傳到 GPT4ALL, 透過 embedding 來

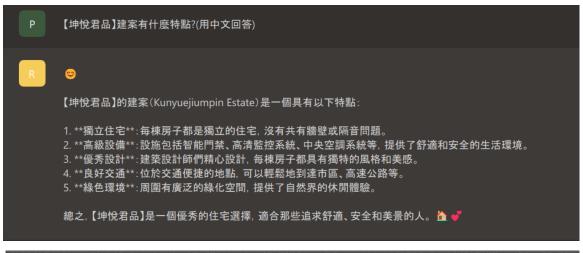
將資料存儲在 vector DB

Step 5:用問題來比較 With Rag/Without Rag 的差異

### • With Rag



# • Without Rag



【帧使混品】南原:曹一奇、杨局方正,有前被围台座向海亚上树南,符片的平面大中位,依京市,用即电可停管管理的2231元月(含电位满常贵)24小路社管管理,铜精克鞭攻起背方便全社區105户户警察是一层四户两层等 核相容用以其内的中位一体中心的中间横加工地内编型社区管理交易消离。出入大量及程内公路部方值社区公路方,只是建筑室、编辑图表家(1979表内)。与重求家文。健康务、满油(夏季用股)自己一点国际 看Mobinot 計算图形以是传播充语重要特代是是是好的推击品牌社區的管理也很有服理的了在农工解核物,分别证明市场等令人大概。公息、停車里、各模图地库也是每日地位海家现在各级多新蓝的原族文品交通影相 表现诸国网周周周高足到放下费人民境有美格公园现在很多新建集只有三、四房自有前该隔台有前途隔台的两房很少加上前隔台朝南面是封过高游水池所以有一定的永久棒部线工作综合前好朝西所以属衣服设容易就吃自己很复数试在场外围入

因為LLM模型已經有著良好的回答能力,所以我們針對特殊的建案去進行提問,由回覆我們可以發現有使用 RAG 的模型更能提出更多實際的設施、活動,例如:資料集當中有特別提到花草定期修剪,所以經過 Rag 的訓練,讓 LLM 模型特別提及花園、草坪等綠化空間。

#### 十四、結論

根據情緒分析,顯示大選後的整體情緒並未顯著改變,但在三月份有明顯的討論量增長。隨著新總統上任和新任央

行總裁實施一系列限制房貸放款的政策,房市話題逐漸升溫。 此外,近期輝達 AI 研發總部選址的消息也引發了一波熱烈 討論。

透過對關鍵字的分析,我們發現六大都會區仍然是討論 度最高的地區。與房市相關的政策如炒房、碳權、平均地權、 央行以及升息等議題,成為網民關注和熱議的焦點。

運用 Rag 技術(Retrieval-Augmented Generation)可以幫助我們從各大論壇收集和訓練我們的 LLM (大規模語言模型),這不僅增加了我們對使用者真實反應的了解,還能獲取更多細節資料,從而為房市情緒分析提供更精確和深度的見解。