



社群媒體分析

專案一

指導教授 黃三益 教授

第13組:

N104020003 王博玄

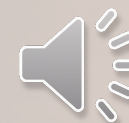
N104020004 葉子豪

N104020015 李宇婕





大綱



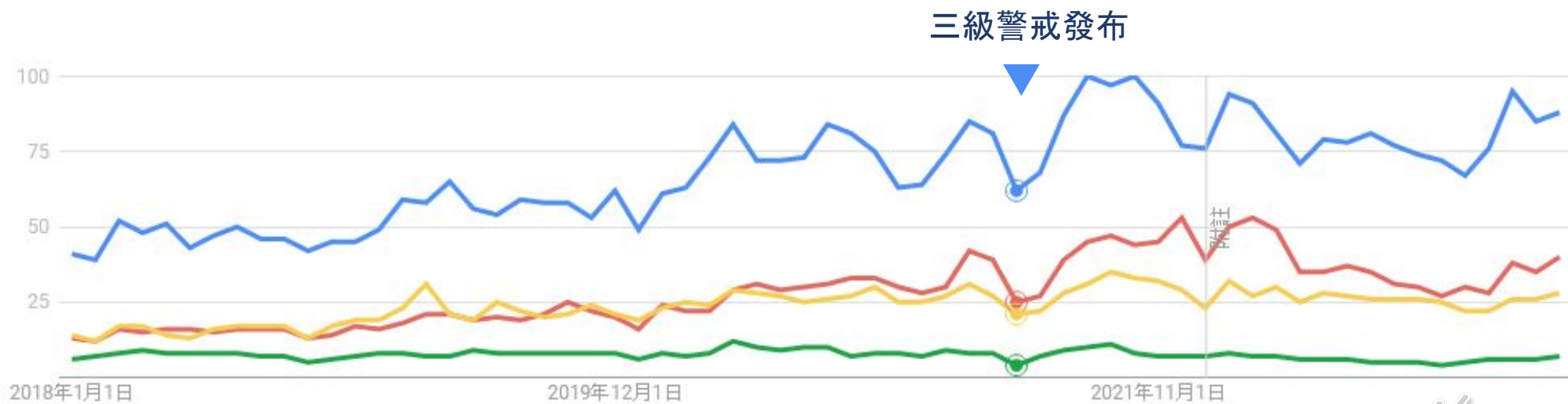
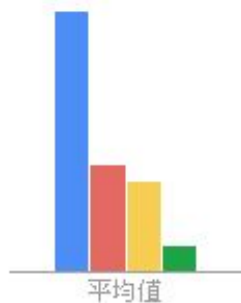


動機及分析目的

／ 依據google trend分析，自2018/1/1(疫情前)至今，「買房」的搜尋熱度穩定爬升，表示買房確實有一定的剛性需求，相關議題如房價、房市近年也時常活躍於新聞版面，而依據物件類型不同，搜尋熱度也不同。

搜尋熱度的趨勢變化

- 買房
- 預售屋
- 中古屋
- 新成屋





動機及分析目的

- 目前購屋大致上可依物件類型大致劃分為3種
- 其中，預售屋無論貸款成數或依施工進度繳款等，均對小資族較為友善，搜尋熱度也最高，故本組以此為研究主題，希可瞭解預售屋相關討論趨勢。

	屋齡短		屋齡長
	預售屋	新成屋	中古屋
貸款成數	80-85% 勝	70-80%	60-70%
優點	勝 <ul style="list-style-type: none">1.按照施工進度繳款2.可配合喜好客變3.全新4.房型選擇多	<ul style="list-style-type: none">1.全新2.可看實際屋況	<ul style="list-style-type: none">1.可看實際屋況2.相較於新屋，同價位能買到較大空間或更好地段
缺點	<ul style="list-style-type: none">1.有建商倒閉風險2.無法立刻入住3.僅能參考樣品屋，有落差風險4.公設比高，目前多為30-35%	<ul style="list-style-type: none">1.頭期款需一次繳清2.公設比高，目前多為30-35%3.為建商餘屋，選擇性少4.無法客變，需預備裝修費	<ul style="list-style-type: none">1.自備款比例較高2.貸款成數較低3.貸款審核時可能有成數上的變動4.需預估一筆修繕費5.可能有仲介費



資料集描述



/// 本組使用課程提供之Tarflow文字分析平台為工具

/// 抓取之資料集設定如下：



- 資料來源:PTT爬蟲
- 看板:home_sale
- 搜尋關鍵字:預售屋
- 排除關鍵字:[賣、[買、[問卷、Re:
- 日期:2021/1/1~2021/12/31
- 資料筆數:1456筆

- 資料來源:PTT爬蟲
- 看板:home_sale
- 搜尋關鍵字:預售屋、房市、房價
- 排除關鍵字:買//、賣//、[買、[賣、[問卷、[買、[賣
- 日期:2021/1/1~2021/12/31
- 資料筆數:5858筆

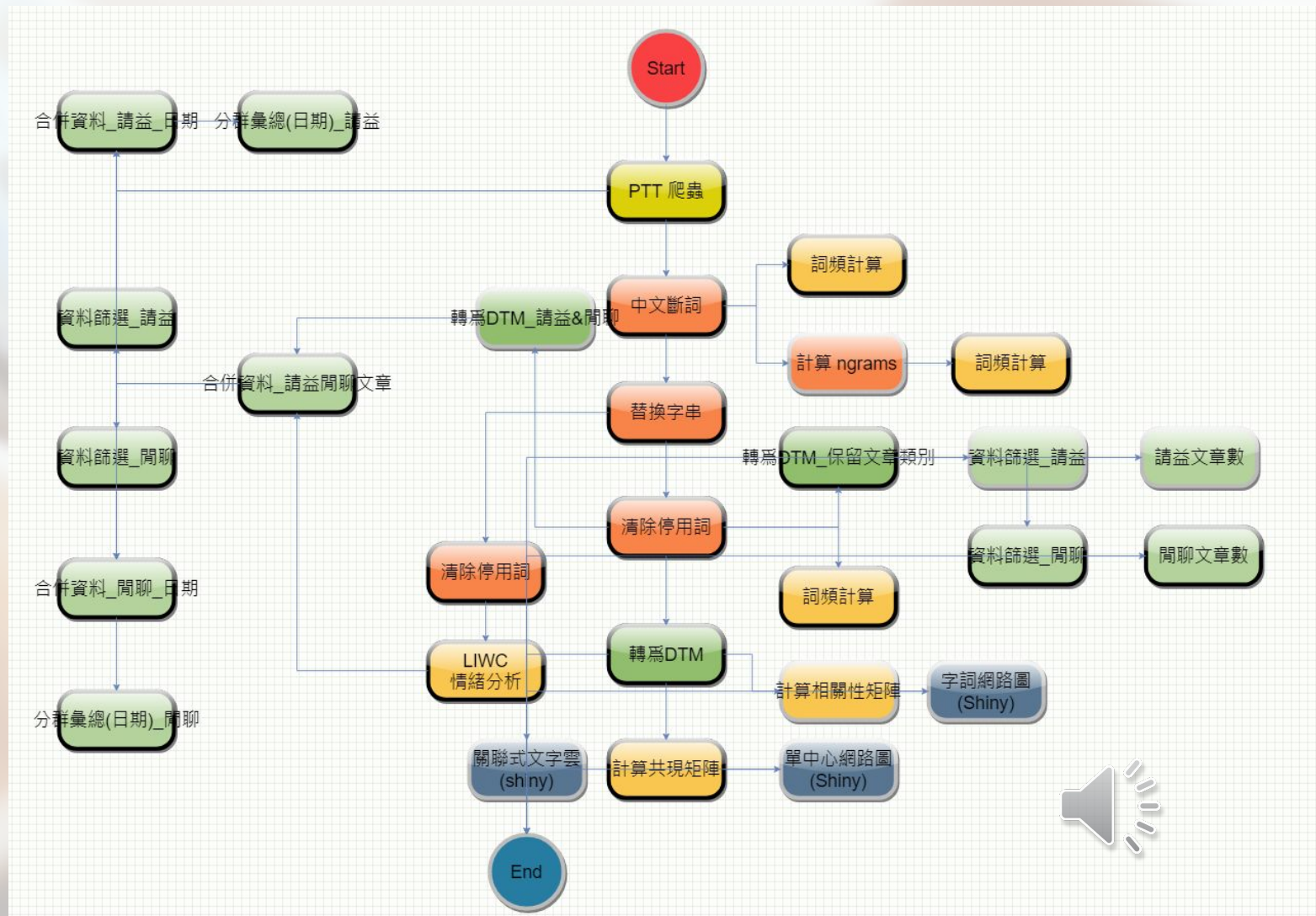
- 資料來源:PTT爬蟲
- 看板:home_sale
- 搜尋關鍵字:預售屋
- 排除關鍵字:買//、賣//、[買、[賣、[問卷、[買、[賣
- 日期:2021/1/1~2021/12/31
- 資料筆數:88853筆





標題之資料分析過程如下：

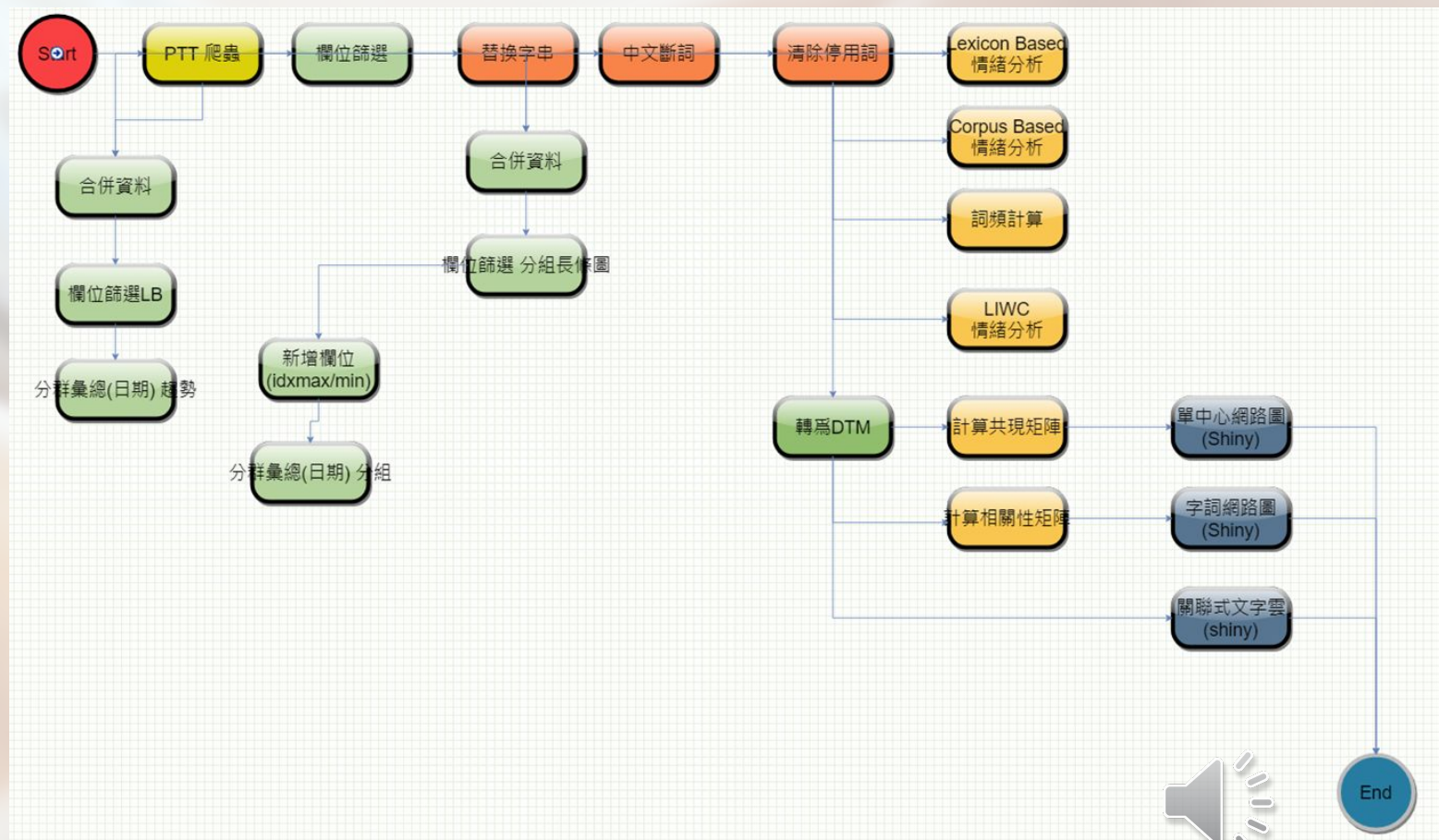
- 將各文章分類、本議題專有名詞(如房地合一稅、實價登錄2.0、使用執照、定型化契約等)加入斷詞，並由n-grams回饋斷詞結果
- 取代特殊字元、處理錯字等
- 自定義停用詞包含問題、內政部、請問、上路等23個
- 透過轉為DTM、詞頻計算、資料篩選、視覺化元件等進行後續關注議題之分析及繪圖





內文之資料分析過程如下：

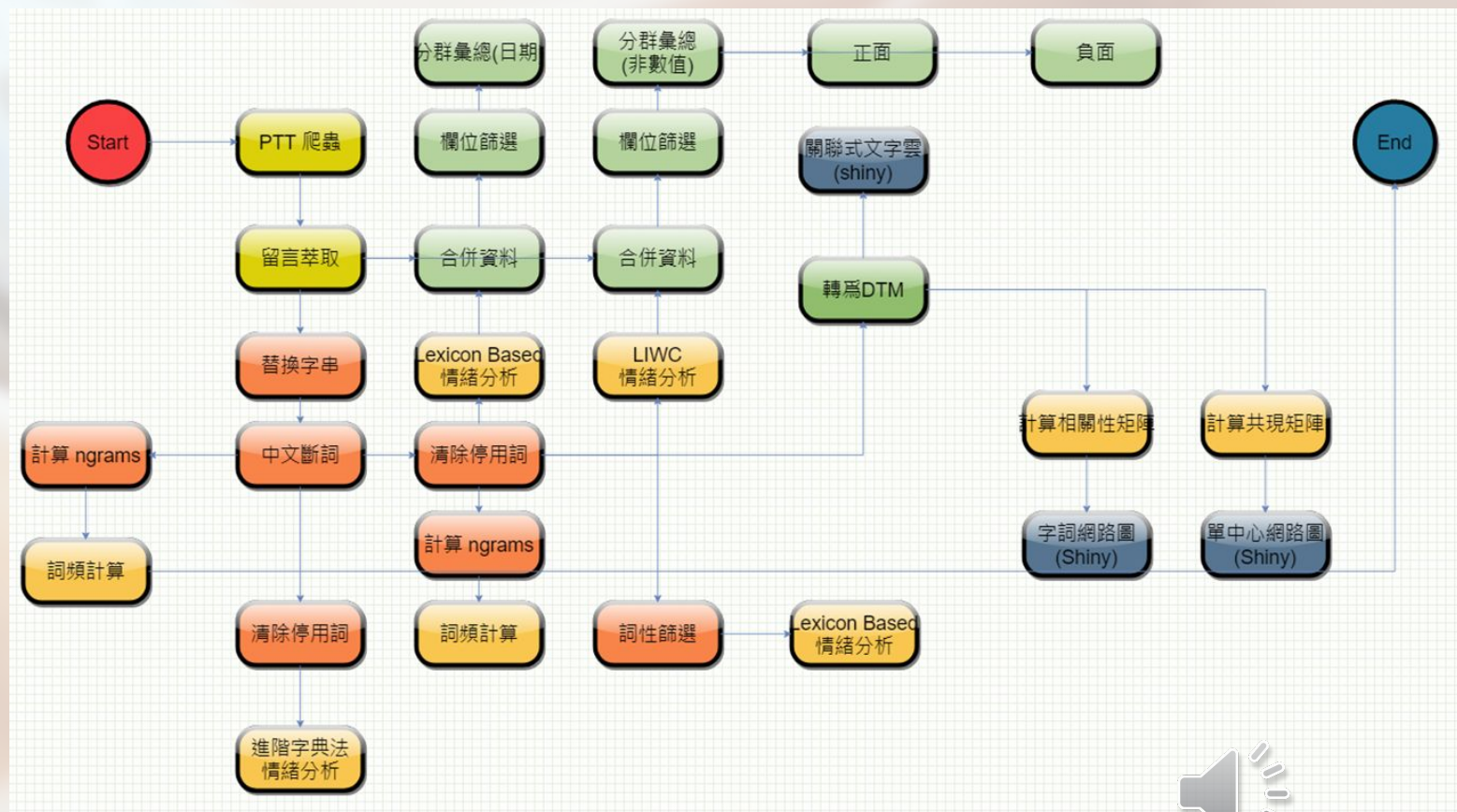
- 篩選內文分析所需欄位
- 移除如網址、特殊字元、取代換行符號等
- 自定義房市政策、房市交易、實價登錄2.0等12個斷詞詞彙
- 自定義清除停用詞「問題」
- 透過轉為DTM、欄位篩選、新增欄位、視覺化元件等進行後續關注議題之分析及繪圖





/// 留言之資料分析過程如下：

- 使用留言萃取功能提取留言
- 移除如網址、特殊字元、取代換行符號、PTT預設載入之社群規定等
- 自定義實價登錄、居住正義、房地合一等22個斷詞詞彙，並由n-grams回饋斷詞結果
- 透過轉為DTM、欄位篩選、合併資料、視覺化元件等進行後續關注議題之分析及繪圖





視覺化分析結果與說明

標題



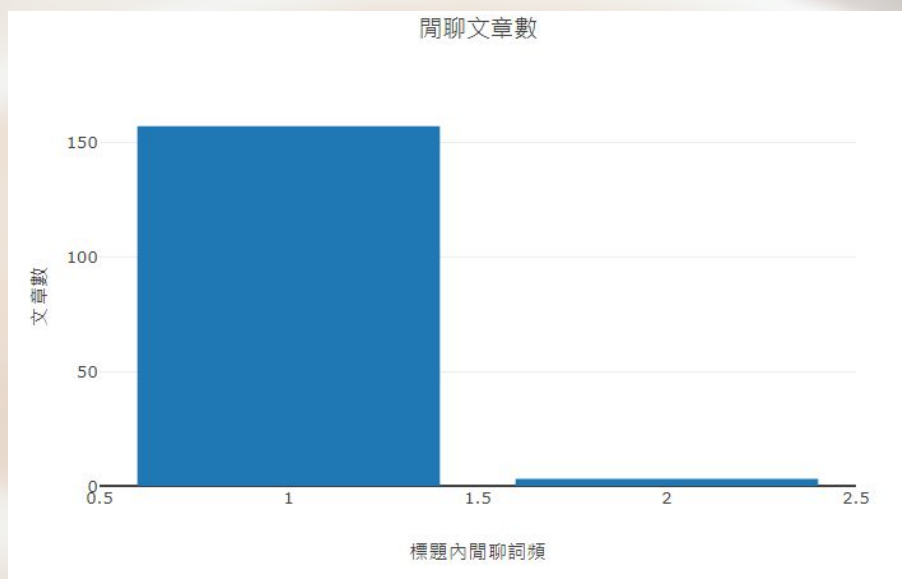
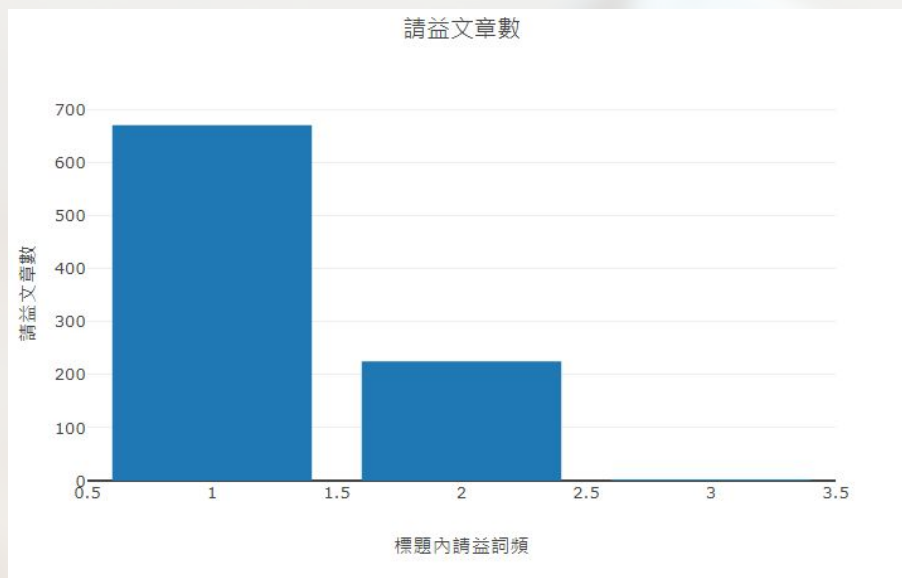
- 清除停用詞後，針對結果進行詞頻分析
- 以格局、建商、房價、等關鍵字最常出現，另有實價登錄、車位、代銷、驗屋、房貸、房地合一稅等議題受到關注。
- 其中也可看出在PTT房屋版上特別受到關注的置產縣市包含台南、北屯、桃園、新竹、新北、林口、雙北、台中等
- 標題之分析關鍵字僅下「預售屋」，然新聞常出現的「房價」並非最頻繁出現之關鍵字，而是「格局」(圖中另有二、兩、三房應皆屬格局)，顯示參與討論的使用者(或可說潛在購買者)可能比起房價，更注重預售屋多樣化的格局選擇。





視覺化分析結果與說明

標題



- PTT具有文章分類之設計，以便使用者快速找尋文章，home-sale版設有「買」、「賣」、「新聞」、「情報」、「請益」、「心得」、「閒聊」、「問卷」之分類，其中買、賣、問卷已於爬蟲時排除
- 其中請益、閒聊較可激起討論，因此選定2分類進行進一步分析(心得多為購屋畢業文)
- 經轉為DTM及資料篩選處理，本組將標題內至少出現1次請益者歸類為請益文章，文章數共計897篇；其中出現3次者多為使用者重複於標題插入文章分類所致
- 閒聊文章定義同上，共計160篇
- 經檢視，請益文章為home-sales版主要大宗文章分類

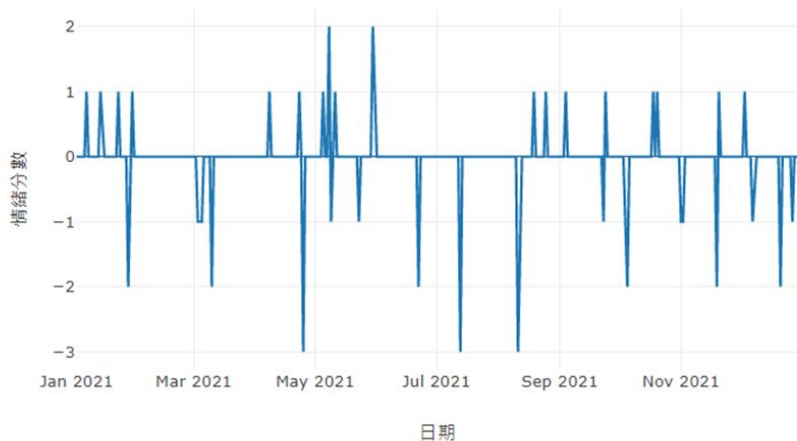




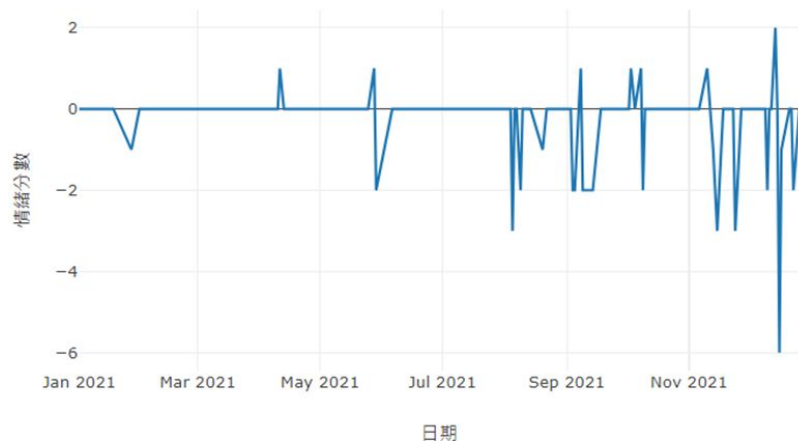
視覺化分析結果與說明

標題

請益文章情緒分數



閒聊文章情緒分數趨勢

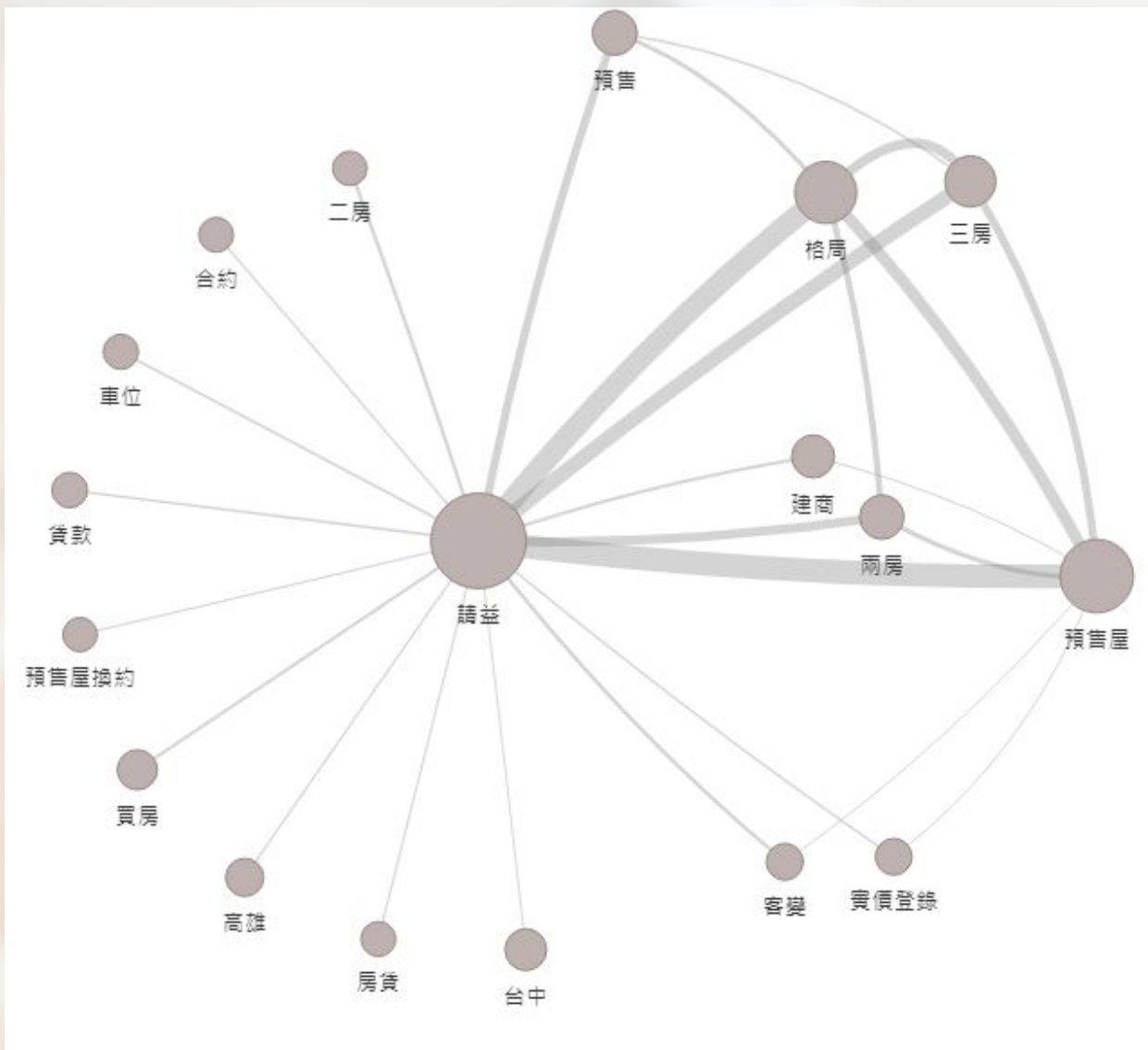


- 經執行LIWC情緒分析，並透過轉為DTM、合併資料、分群彙總等處理，計算兩類文章之情緒分數
- 可見請益及閒聊文章之情緒分數相對平穩，可能與其多著重於功能性分享與討論有關
- 閒聊文章在12/16有一波情緒分數的急遽下降，但返回檢視資料，因預售屋在搜尋日期期間發生「禁止預售屋轉約」之政策頒布，而「禁止」一詞在LIWC字典中被分類為負面詞彙，才導致當天情緒分數急遽下滑；「禁止轉約」本身應屬中性的詞彙，雖已透過停用「禁止轉約」一詞來平衡，但網友使用的排列組合多樣，例如12/16當日網友使用的詞彙包含「禁止預售屋轉約」、「禁止轉約」、「禁止轉讓」等，為免過大範圍停用而導致流失資訊，暫且未將以上詞彙停用。



視覺化分析結果與說明

標題



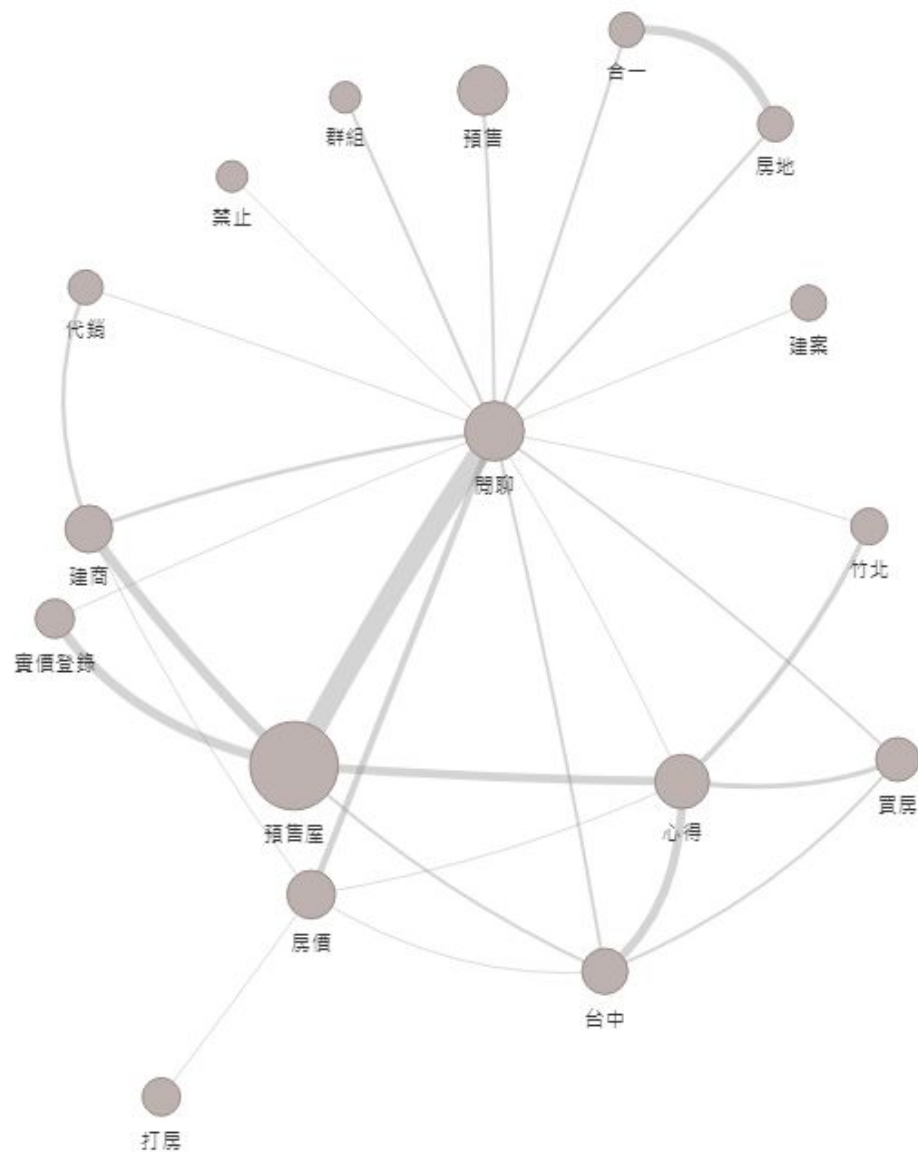
- 以網路節點:20、關聯強度:0.7繪製單中心網絡圖
- 當選取「請益」為詞類中心時，可見請益文章中二房、格局、車位、貸款、客變、建商等為與其相關性較高的討論關鍵字，議題具體對焦。





視覺化分析結果與說明

標題



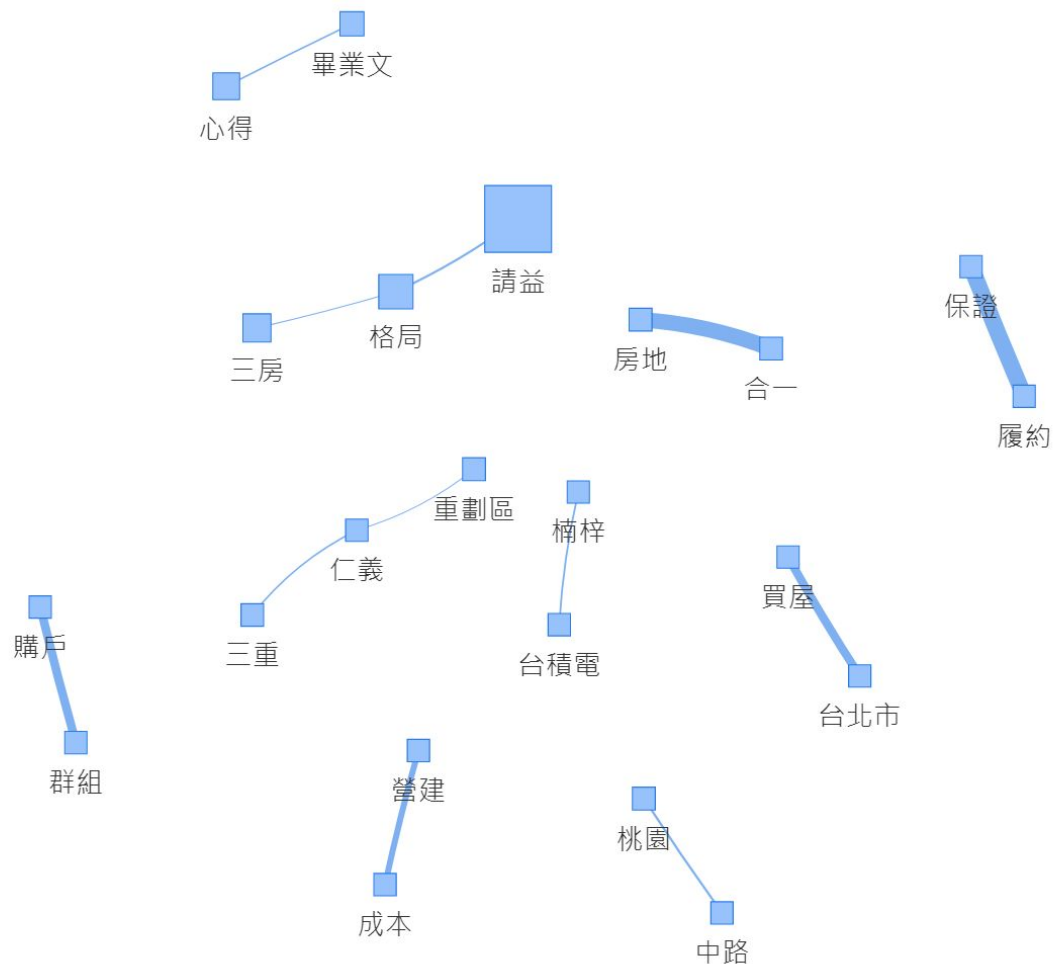
- 以網路節點:20、關聯強度:0.7繪製單中心網絡圖
- 當選取「閒聊」為詞類中心時，可見閒聊文章中房價、實價登錄、建商、房地合一等為與其相關性較高的討論關鍵字，議題範圍就相對廣泛發散





視覺化分析結果與說明

標題



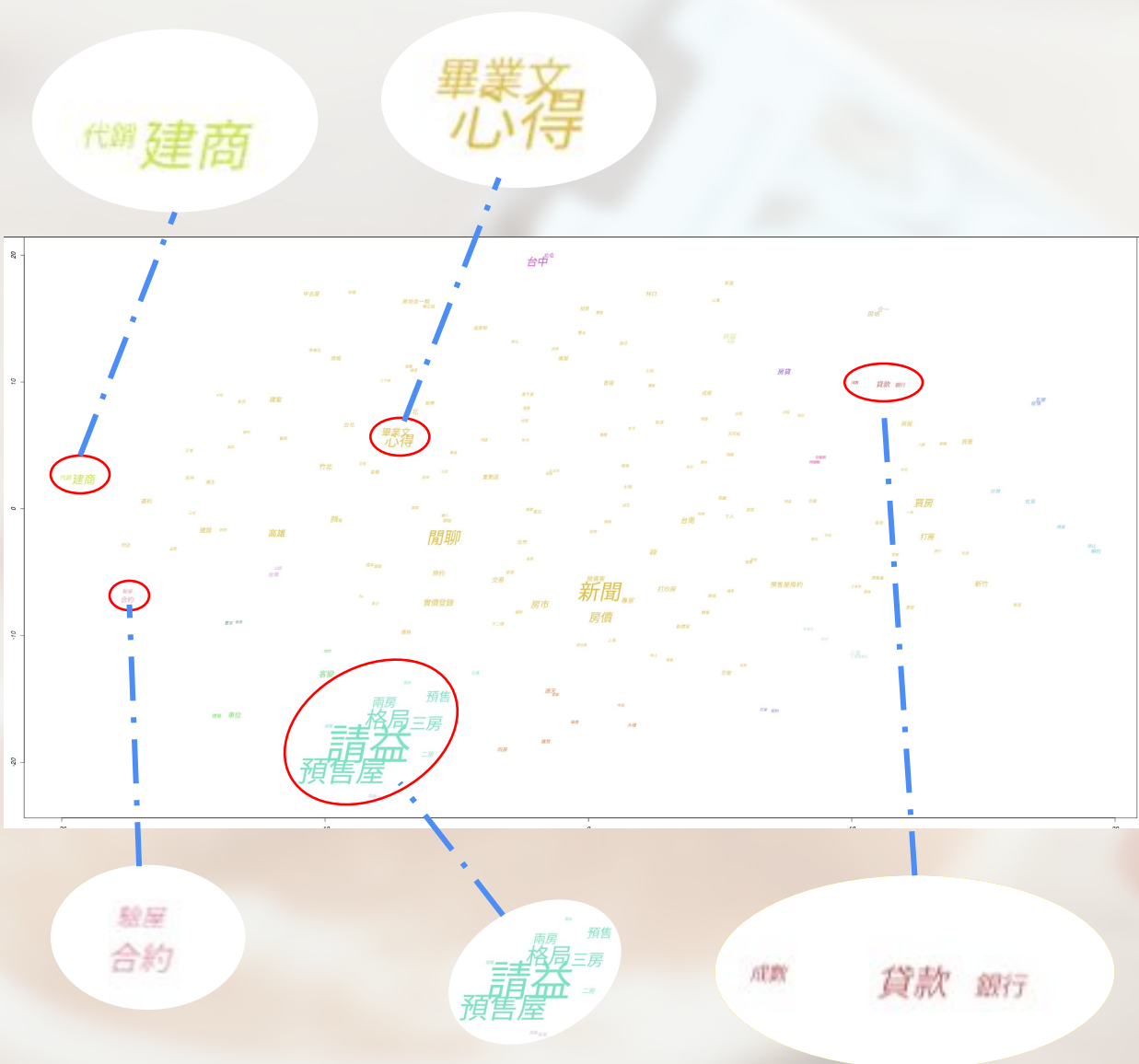
- 以關聯強度0.5繪製字詞網絡圖
- 可見請益確實常伴隨格局相關關鍵字、有趣的是三重、仁義重劃區、楠梓台積電、桃園中路重劃區等均有相當強關聯
- 此分析可指出使用者較為關注的買房地點，重劃區可能是使用者較為偏愛或有興趣的置產地點



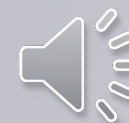


視覺化分析結果與說明

標題



- 可見幾個比較明顯的討論議題，可分別歸類為格局、建商代銷、銀行貸款、買屋心得等相關討論主題。

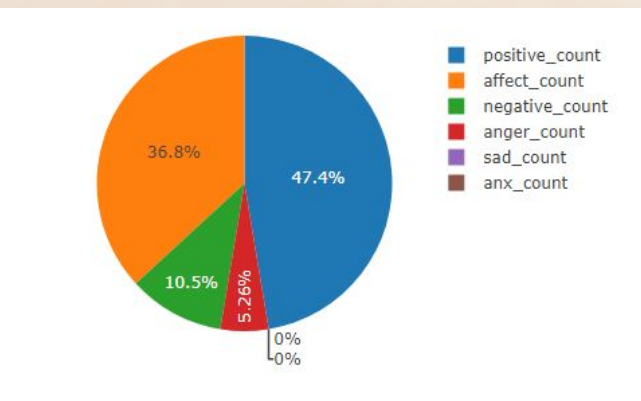
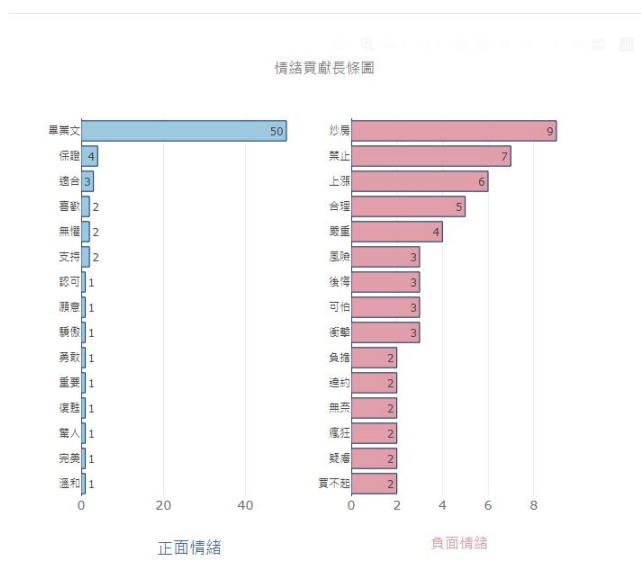




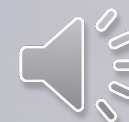
視覺化分析結果與說明

標題

情緒分佈



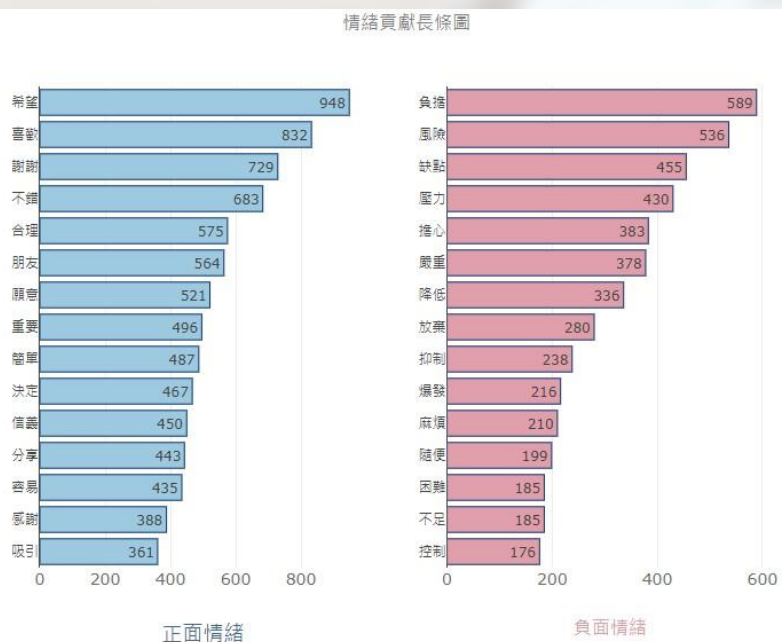
- 在情緒分析當中，考量關鍵字為預售屋，故將買貴、合理(通常語境上詢問合理時，通常都是在懷疑不合理，故分為負面)、房價漲、價漲、漲、漲多、上漲、炒房、買不起添加為負面詞彙；將跌、下跌、房價跌、畢業文(於此看板，畢業文表示作者已買到房)等列為正面詞彙。
- 總體而言，看似正面多於負面，然多為畢業文所貢獻；扣除畢業文，負面情緒將佔比較多，且可觀察到炒房、房價上漲等仍是使用者的負面觀感來源。





視覺化分析結果與說明

內文



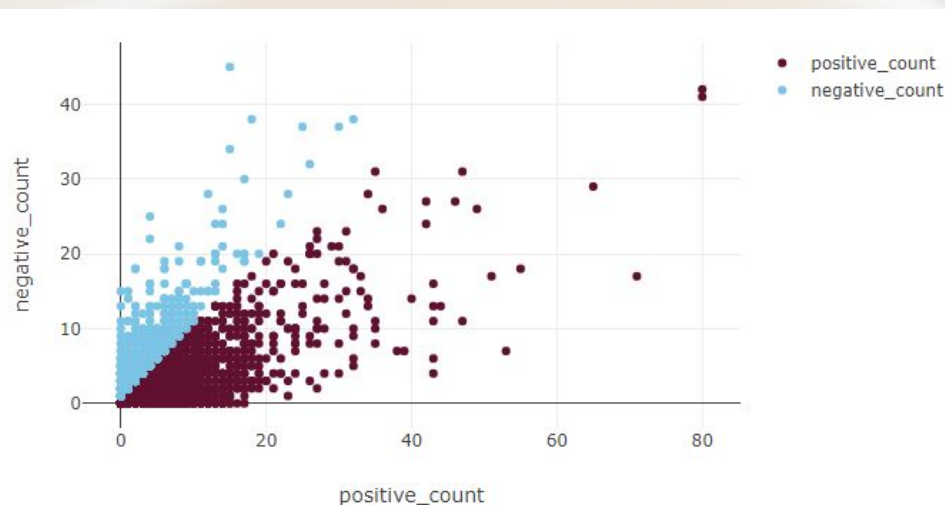
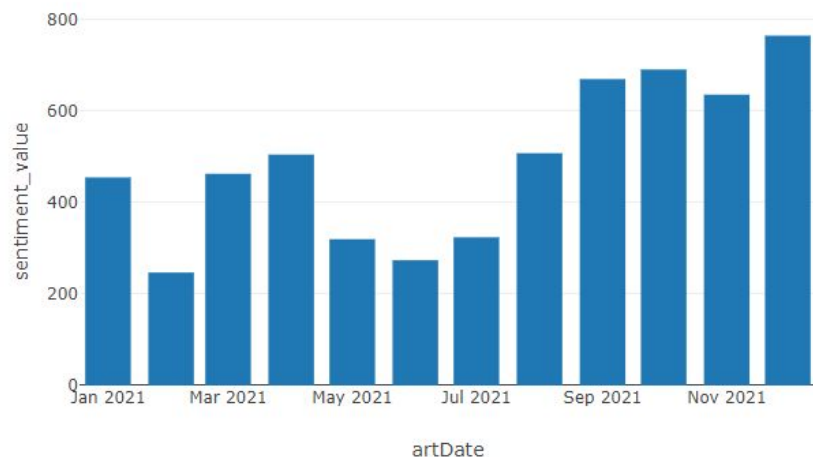
- 執行Lexicon Based情緒分析，可見內文的正面情緒高於負面情緒，但有些詞彙雖被字典歸類於正面，但實際語境應用上可能較難區分正負面，如「希望」於文章中多用於表示使用者購屋時的期待、「合理」於內文中多使用於開啟或接續話題(如你們覺得房價合理嗎、討論某樣事物合理性等)、朋友則可能用於分享周遭親友購屋經驗等
- 而負面情緒第一名「負擔」，仍反映房屋版使用者對於房價、貸款等的熱切討論及擔憂情緒



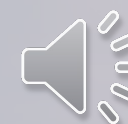


視覺化分析結果與說明

內文



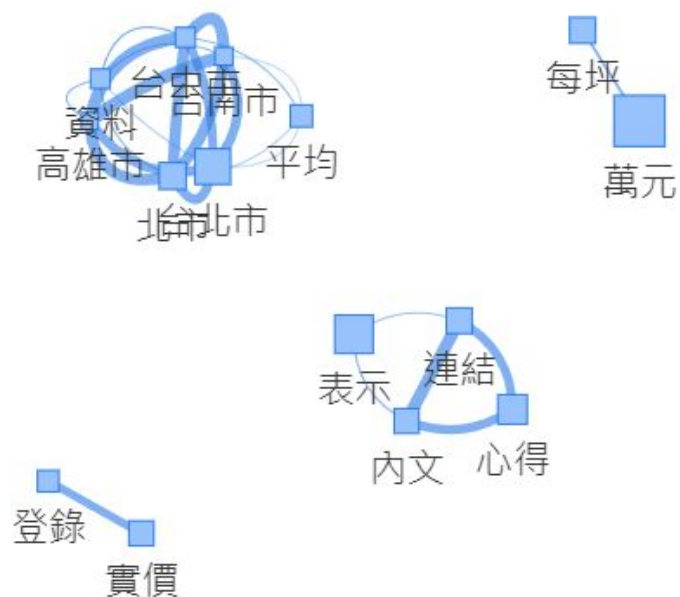
- 觀察Lexicon Based情緒分數與時間之趨勢，2021年受到疫情持續影響，可見在年初的情緒分數較低，可能為討論量下滑或使用者對房市較悲觀。
- 7月以後可見情緒分數逐漸走高，表示在疫情緩和後，使用者對房市的態度可能有所好轉，逐漸趨於正向。
- 整體而言，房市相關的討論內容仍有正有負，但可見持極端正面態度的文章較負面態度多，整體使用者對房市應仍是抱持樂觀看法。





視覺化分析結果與說明

內文



- 以關聯強度0.55繪製字詞網絡圖
- 高雄、台中、台南、台北市等具有相當關聯，表示使用者在討論購屋選擇時不會只看一個地區，且推測六都置產的討論仍相對其他縣市熱烈
- 每坪單價、實際賣出價格仍是使用者討論重點
- 注重其他人的分享



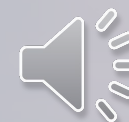


視覺化分析結果與說明

內文



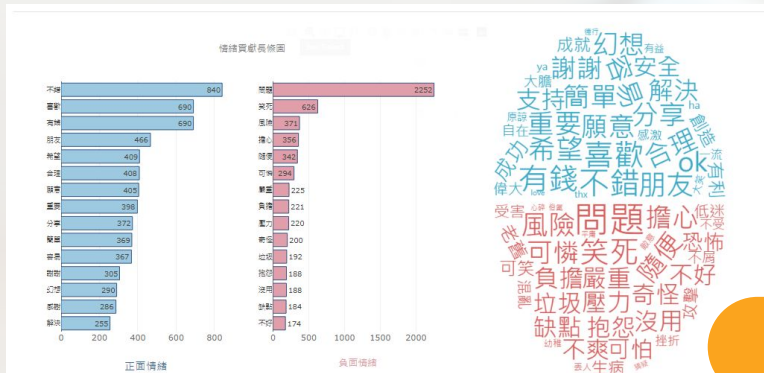
- 可見價格無疑是購屋者最關心的條件，如房價、漲幅、租金、上漲等
- 其次也可見貸款相關討論，如利率、收入、通膨、負擔等
- 政策也是一大類議題，如投資客、炒作、房地產、打房等財經討論
- 物件類型如成屋、預售屋、中古屋等
- 地區相關議題則包含縣市、建案、代銷等



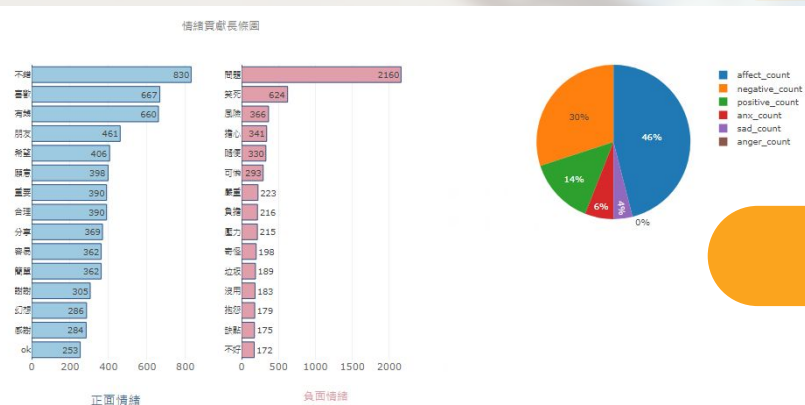


視覺化分析結果與說明

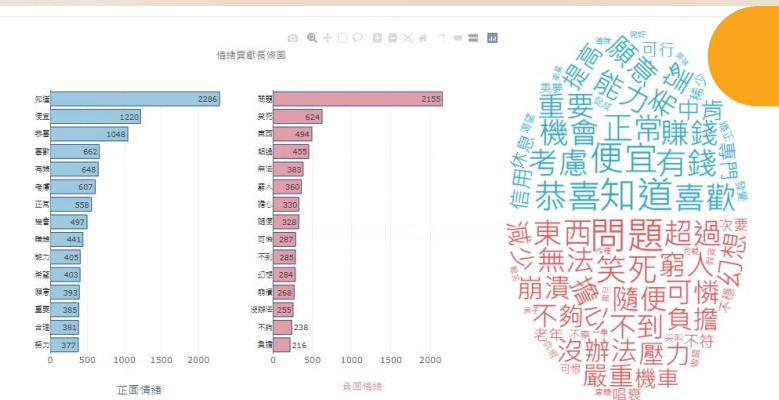
留言



Lexicon Based

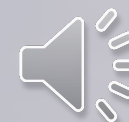


LIWC



進階字典法

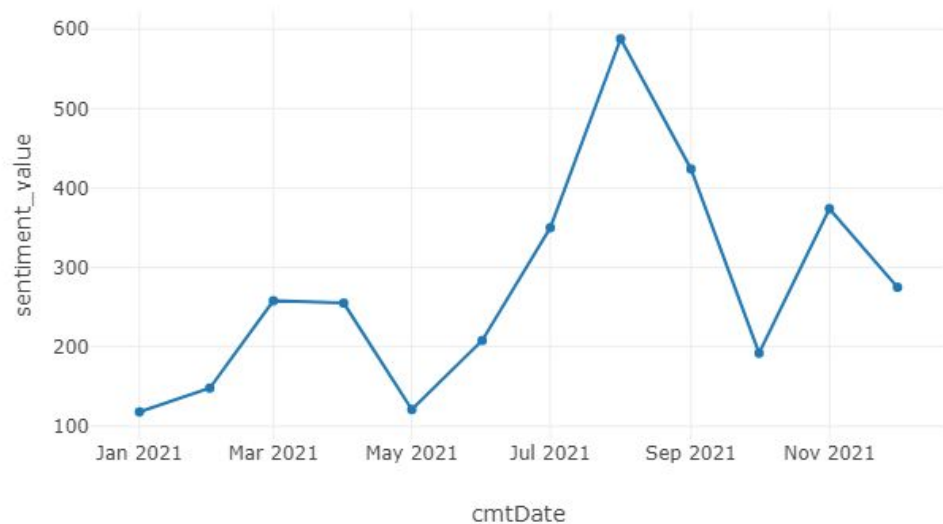
- 留言可以說是最強烈展現情緒的地方，因此本組利用三種情緒分析法分析留言
- 可見三種分析方法負面情緒均類似，而正面情緒方面，進階字典法將「知道」、「便宜」列為正面詞彙。
- 三種分析方法中的正面情緒詞彙對於本組研究議題，都有尚待商榷的語境，如「幻想」、「簡單」、「希望」等





視覺化分析結果與說明

留言



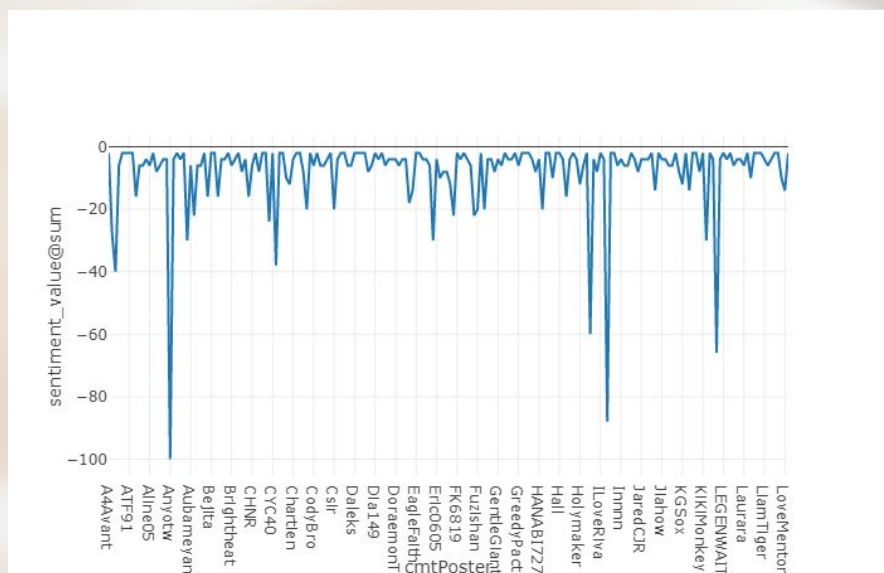
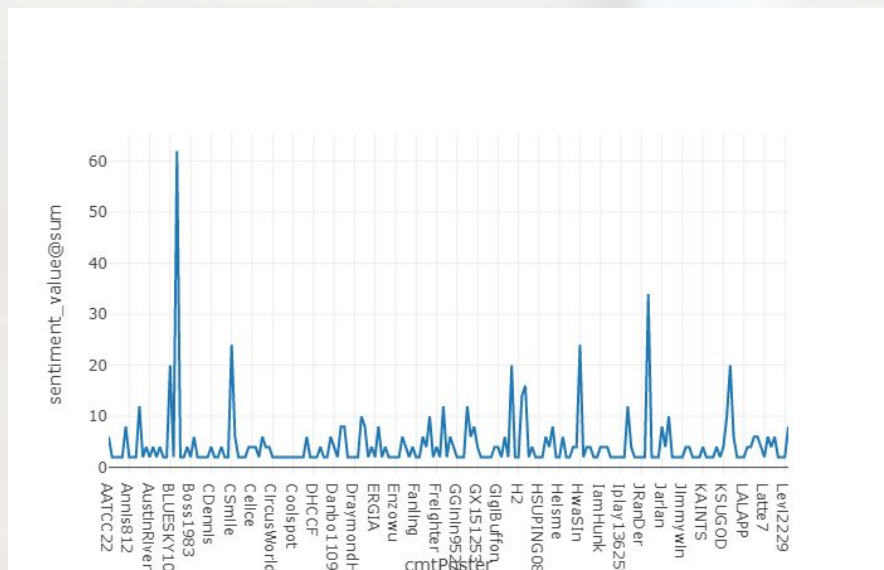
- 分析留言的情緒分數與時間趨勢，可見2021年5月時有一段明顯下滑，推測可能是因為當時疫情爆發，使用者較多關注於其他議題，也較無法出門，使預售屋討論熱度下降
- 2021年8月起，情緒分數明顯上揚，推測可能為三級警戒解除、疫情趨緩，故預售屋相關討論再次回溫



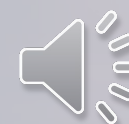


視覺化分析結果與說明

留言

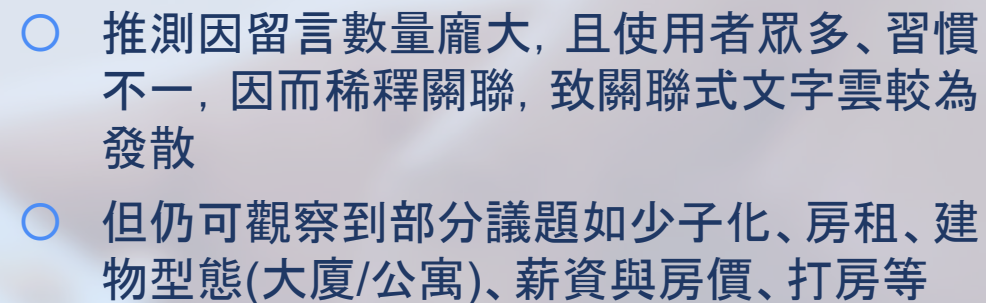


- 針對留言作者進行分析，可見帳號「BRIANERIC2」整體情緒分數相當高，可能表示此作者對預售屋抱持極正面態度
- 凡之，作者「Anyotw」整體情緒分數則相當低，可能表示此作者對預售屋相關議題的反應較為負面。





留言





結論

預售屋相關討論文章，以「格局」討論最為熱門；且重劃區為備受關注地點

1

使用者對於房市的態度及討論熱度會受到疫情影響；疫情趨緩後仍多持樂觀態度

2

留言之情緒語境較為複雜，且可能受總數量高之影響，關注議題亦較為發散

3

