社群媒體分析期中專案報告

題目：

雙11購物於PTT討論版之情緒分析

指導老師：黃三益教授

第三組 郭耿耀N094020005

周昆穎N094220007

李昀瑾N094020022

王莉筑N094220029

# 大　綱

[大　綱 2](#_Toc88338043)

[一、動機和分析目的 3](#_Toc88338044)

[二、資料集的描述 3](#_Toc88338045)

[三、資料的分析過程 6](#_Toc88338046)

[四、視覺化的分析結果與解釋 6](#_Toc88338047)

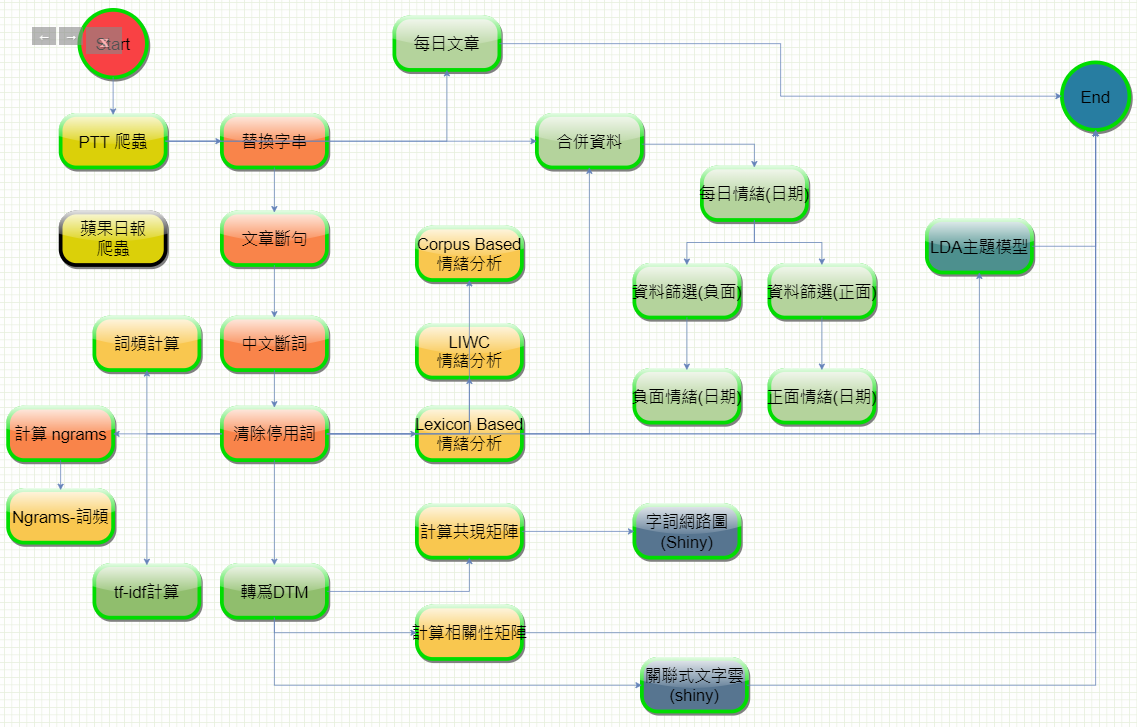
[五、結論 17](#_Toc88338048)

## **一、動機和分析目的**

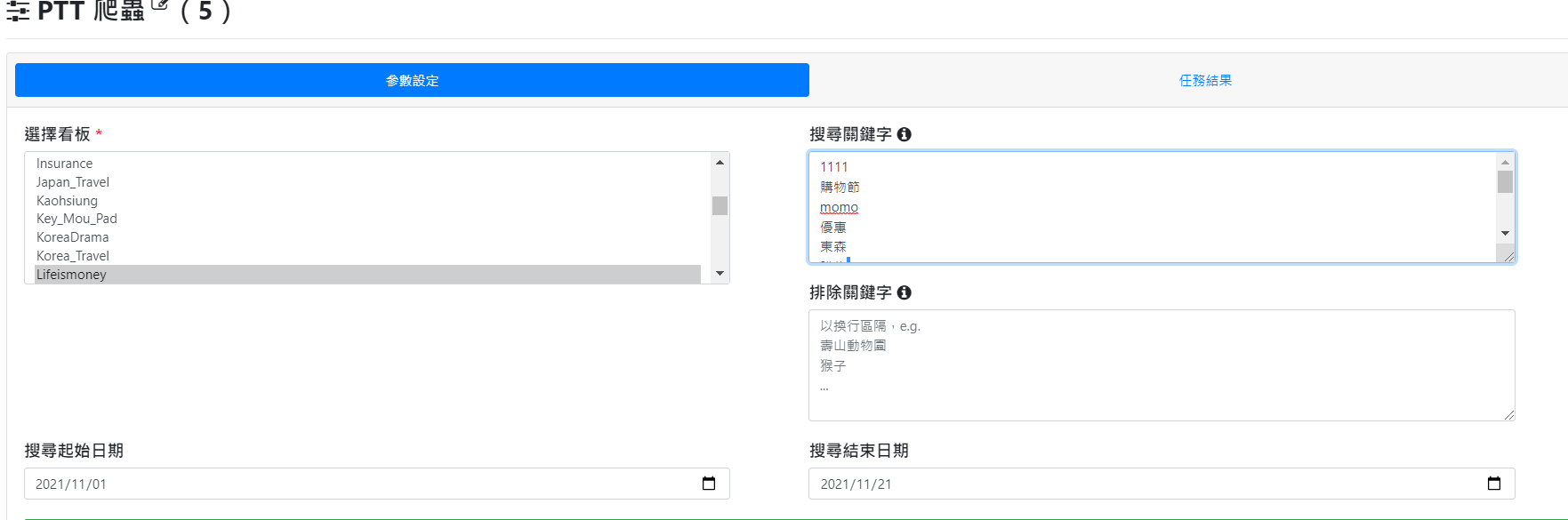
每年第四季是零售業者的重要季度，除了傳統的聖誕節及跨年之外，實體百貨周年慶及網路大型活動雙11和雙12 也都相繼這一季度推出各項優惠，所以如何搶攻消費者荷包會是很重要的一環。當零售業者在行銷產品時，常常會透過各種刺激購買的手法，設法讓消費者產生難以抗拒的衝動進而購買，卻忽略此作法是否可能引來消費者的負面情緒，因此探討消費者購物後的情緒分析的做法具有一定的參考價值。

**二、資料集的描述**

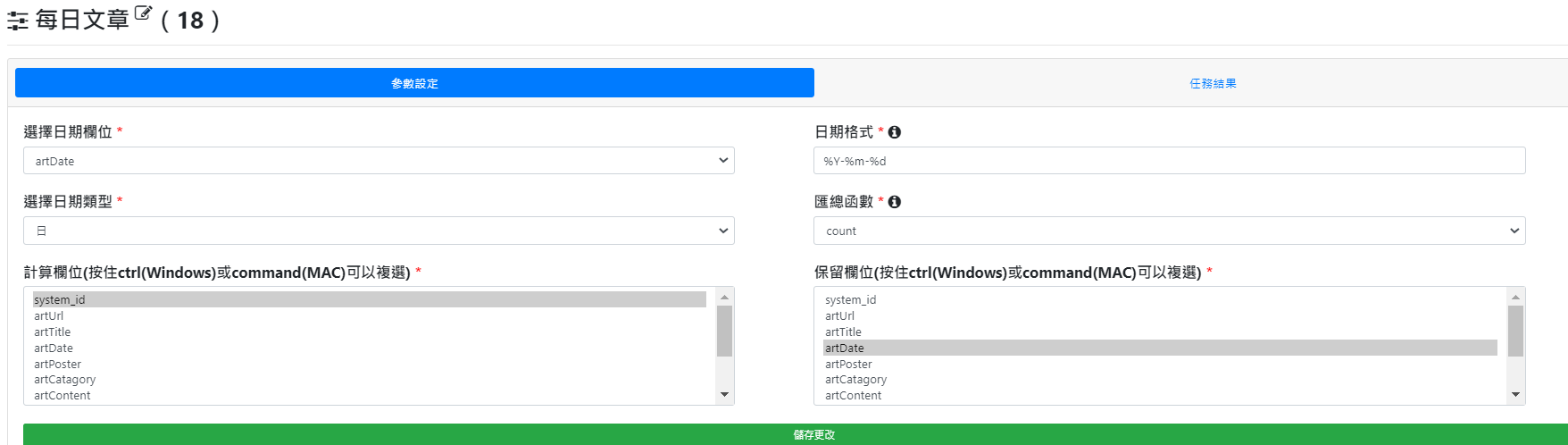
因本組欲探討雙11購物的情緒分析，從工作流程平台既有資料來源爬取PTT於2021.11.01~2021.11.22區間:Bytogether、Lifeismoney兩個有關購物討論版，找出與雙11購物節的相關文章。另蘋果日報吃喝玩樂版爬蟲結果因資料量少，所以就不列入合併分析。爬蟲結果總計取得1,002筆的資料筆數。

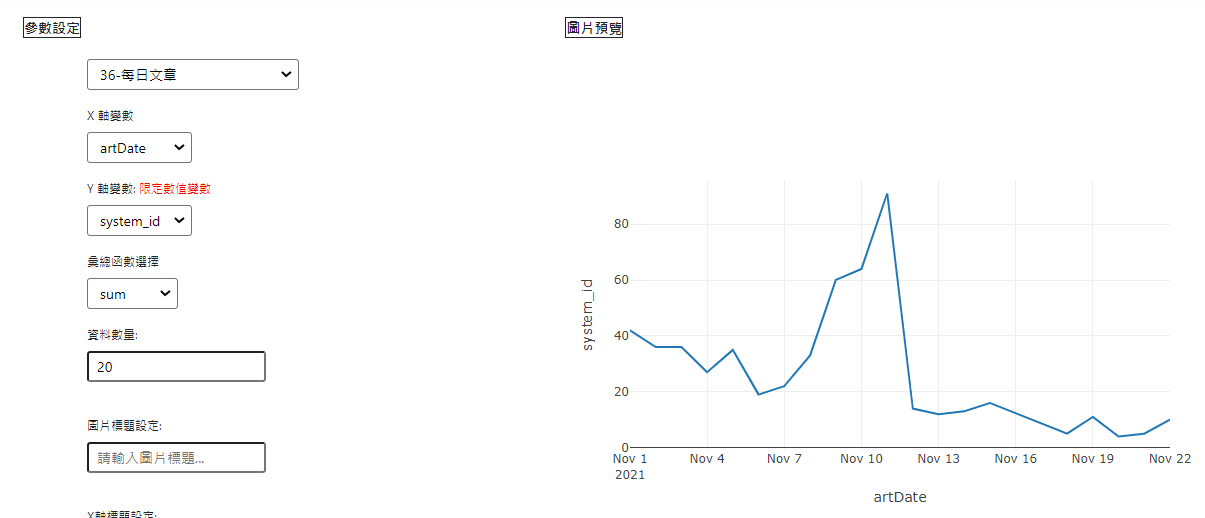
整體流程架構圖

資料平台:**PTT(Bytogether、Lifeismoney)**兩個討論版

關鍵字:111/購物節/momo/優惠/東森購物/momo/pchome/樂天博客來/松果…等知名購物平台及購物相關字詞

從PTT兩個討論版共取得1002筆數資料

PTT每日文章統計

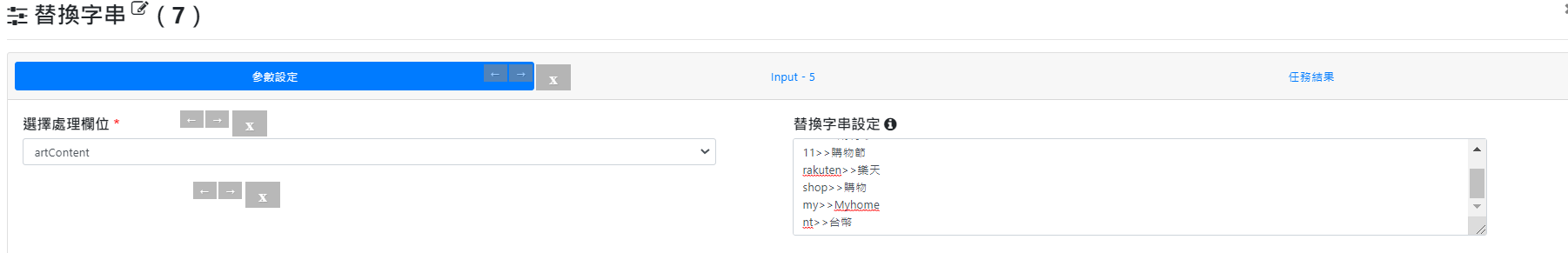
從儀表板產出PTT每日文章，以11/11為討論為最高

**三、資料清理過程**

替換字詞: 為更精準分析將以下字詞作替換，如:

1111>>購物節/111>>購物節/11>>購物節/rakuten>>樂天/shop>>購物

my>>Myhome/nt>>台幣



**中文斷詞:**

將前幾大電商平台詞彙權重提高:momo 1000/pchome 1000/樂天 1000/博客來 1000/松果1000/my 1000/1111 100/優惠 100/東森 1000/購物 100/購物節 100

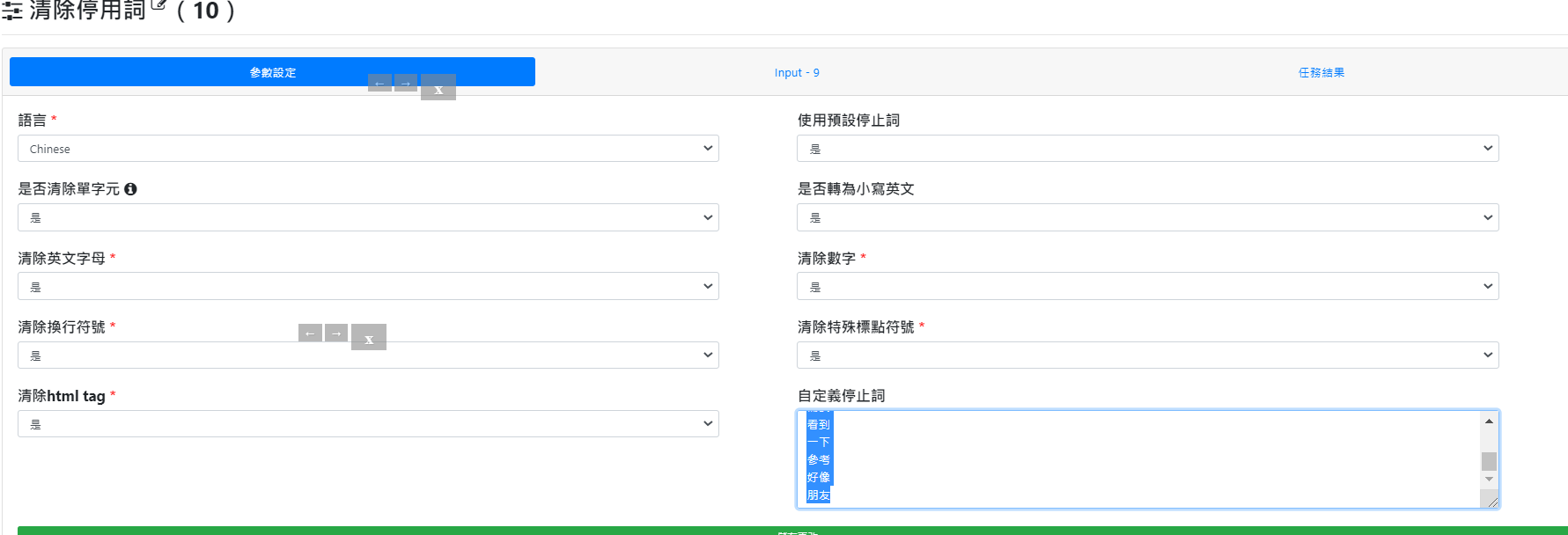


**處理結果分析:**

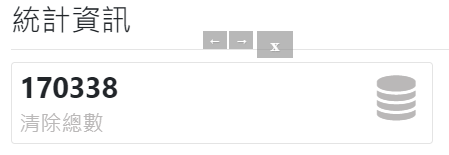
本組欲將熱門平台做統計，但並沒看到關於電商平台討論的熱門度結果(如momo、東森、pchome)等字詞，經抽樣原始文章，發現網友習慣將直接優惠連結直接PO文，及產品內容，較少提到產品是位於哪個平台上。

**參數設定**:

清除停用詞:手動自訂約30多筆，手動將網址及跟內容無關自詞做過濾，並啟用清除html lag\*



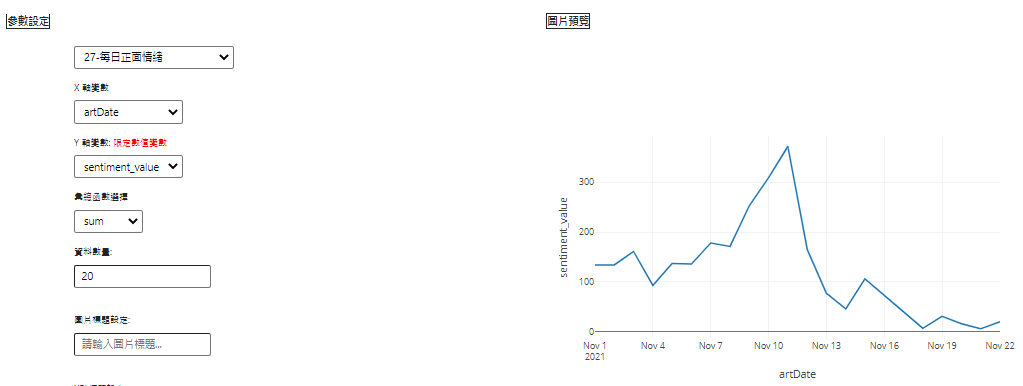
處理結果:共清除170,338筆



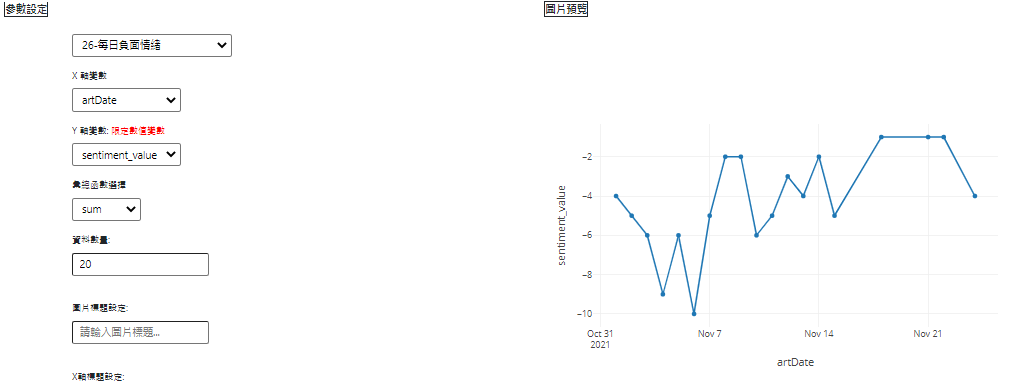
**四、視覺化的分析結果與解釋**

**視覺化儀表板:**

**每日情緒總和: 以11/11為正面最高峰**



**每日負面情緒曲線圖:負面自11/11後高低起伏，後面再進行原因分析**



**計算Ngrams:**

首先將爬蟲結果進行資料前處理，文章包含許多許多英文、數字與表點符號，為了實作方便，將除了繁體中文字以外的字用正規表達式去除後計算NGrams，進行正負面情緒分析詞彙，先設定n=2，找出2個較常共同出現的詞彙，較常出現的詞彙包含付款方式、原價買價等。



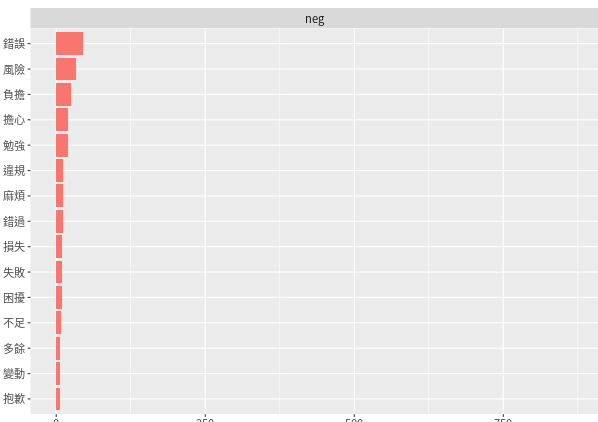


Ngrams文字雲

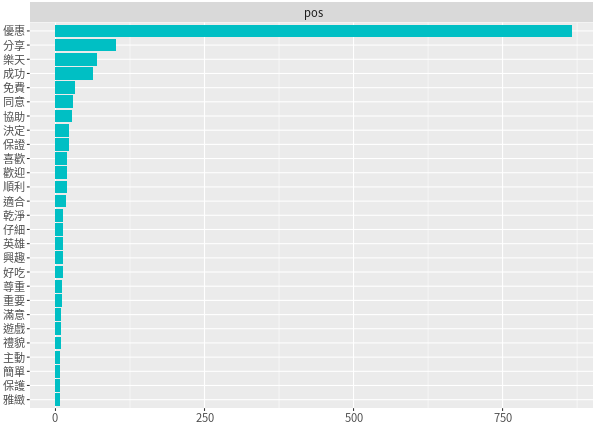
TF-IDF計算結果:

接下來將TF-IDF計算結果進行Lexicon Based情緒分析。其中負面情緒以禁止、錯誤的詞彙出現次數最多，風險、負擔次之。正面情緒以優惠的詞彙出現次數最多，顯現零售業者在雙11節慶推出的優惠價格最受消費者關注。而在分享、免費方面也同樣關注度較高。

Lexicon Based情緒分析:負面情緒貢獻長條圖



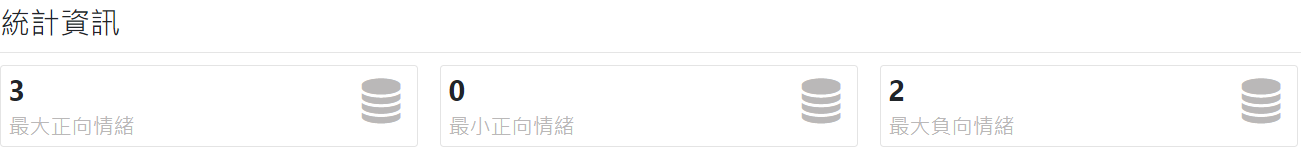
Lexicon Based情緒分析:正面情緒貢獻長條圖



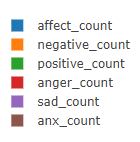


視覺化文字雲

透過LIWC情緒分析結果，取得3個正向、2個負向情緒，及透過儀表板產出Lexcion-Based情緒分析圓餅圖

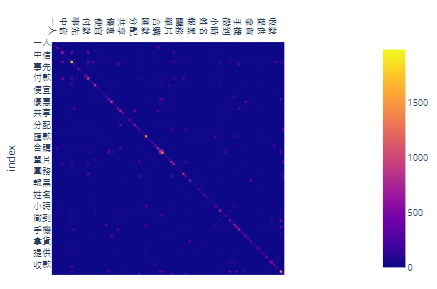






接著計算共現矩陣的目的是要取得2個詞彙共同出現在一篇文章的頻率。

**共現矩陣共200字**

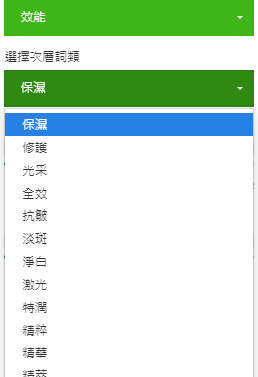


字詞網路圖:

接下來把計算相關性矩陣的結果進行文詞網路圖的詞彙分類，從計算相關性矩陣取得詞彙csv資料檔，並進行分類後匯入字詞網路圖，可於字詞網路圖選定詞類主題為中心，從文集中選出相關性最高的關鍵詞，並對它們進行社會網絡分析。

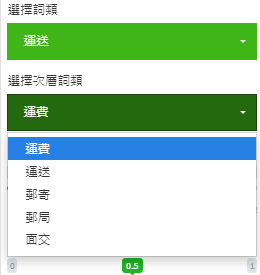


將自選擇詞類及次層詞類，將詞類共處理成12種類別:如產品、消費、效能(功能)、優惠、國家、運費…等類別之熱門詞彙。



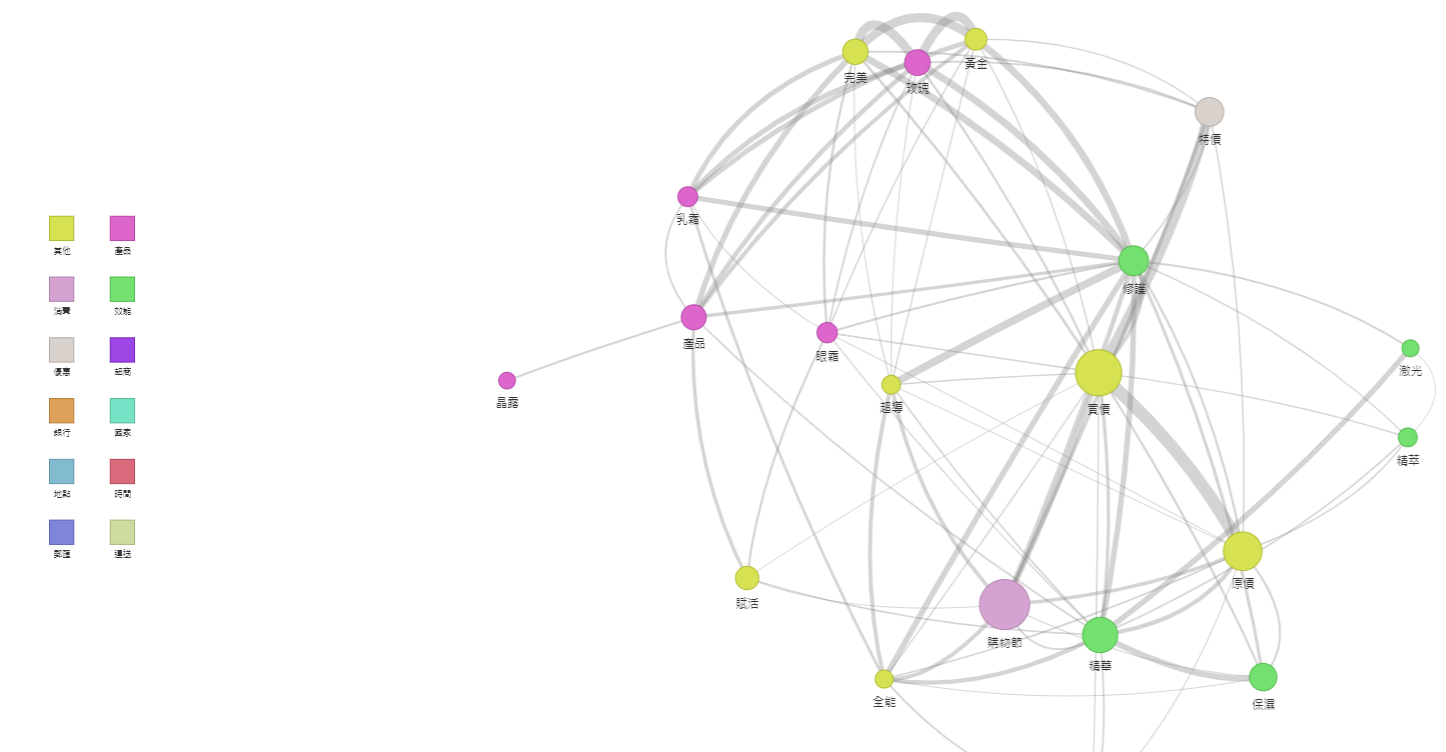


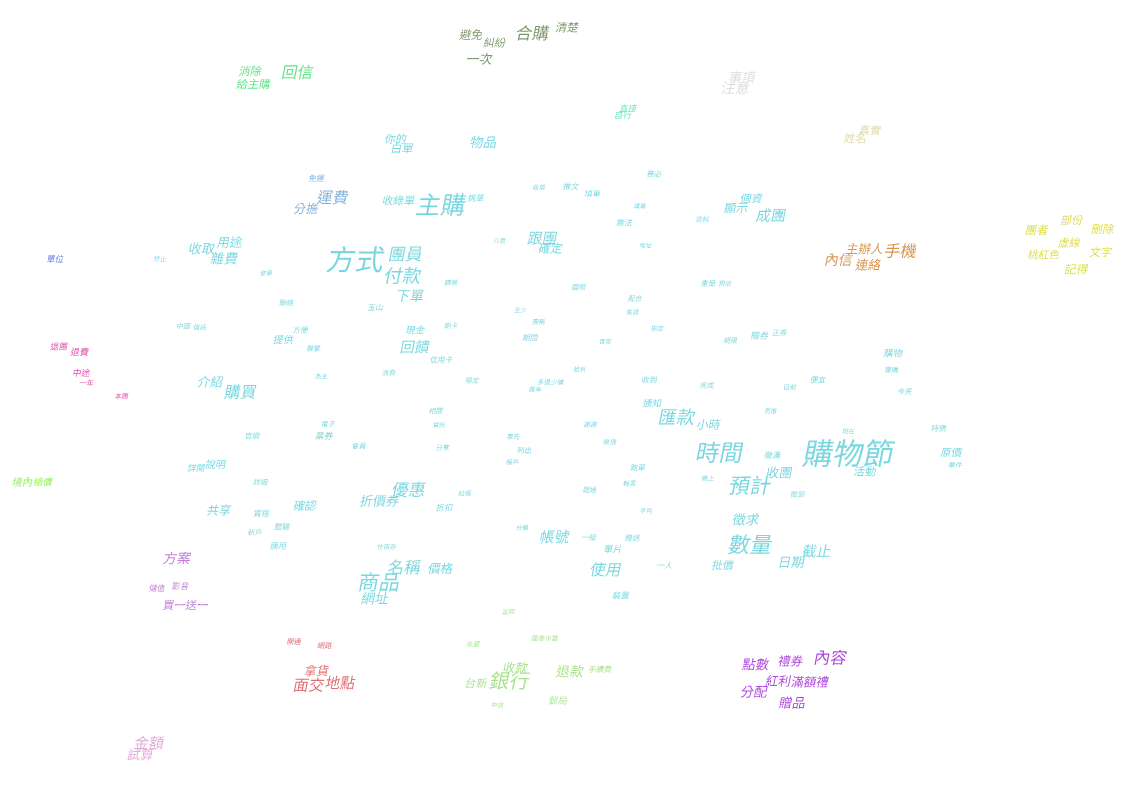




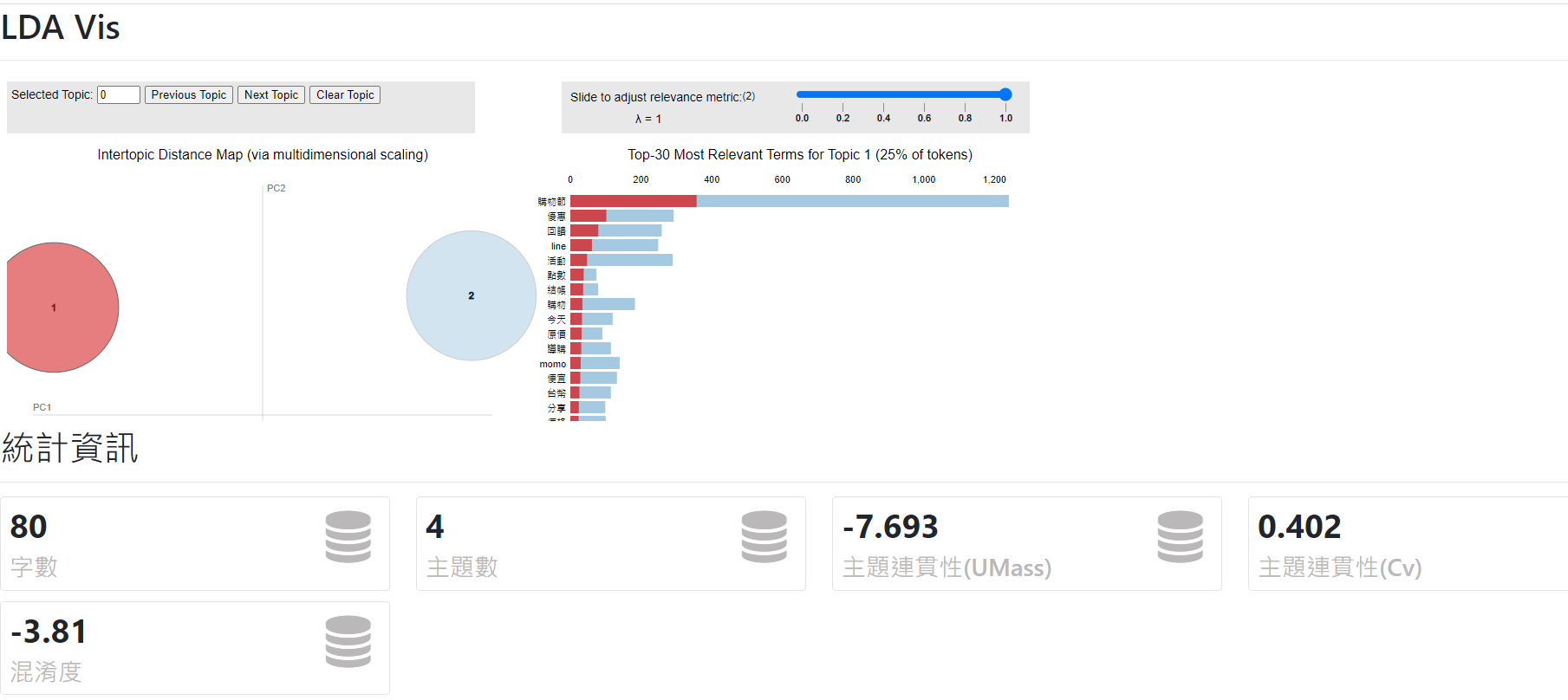


以分類方式產出字詞網路圖



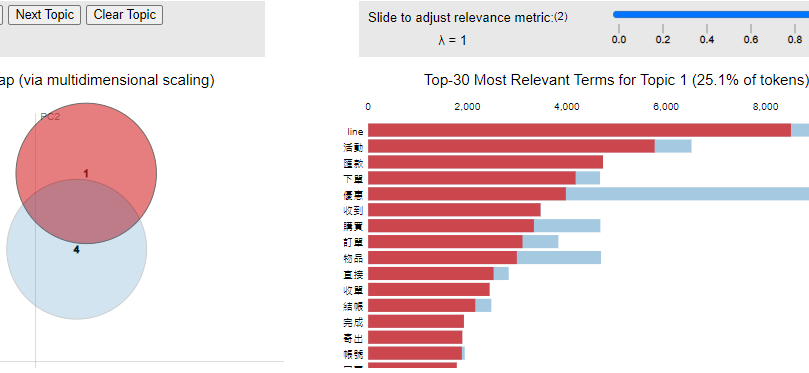
進行關聯式文字雲分類，將具有關聯的詞彙聚集分類在一起。

產出LDA主題模型:設定4個主題數做分類



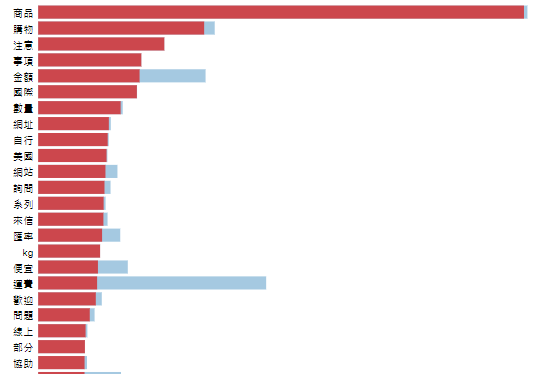
第一主題:(活動/行動類型)

選擇關鍵字:匯款、下單、訂單、購買



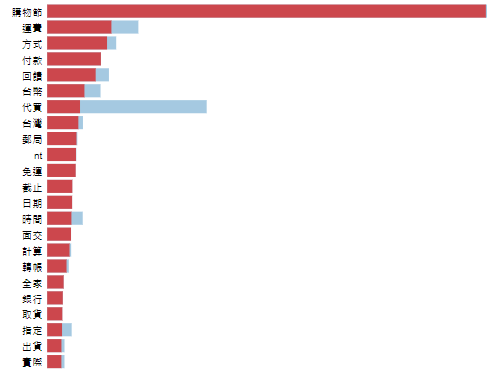
第二類主題:(購物商品相關)

關鍵字:商品、金額、數量、購物



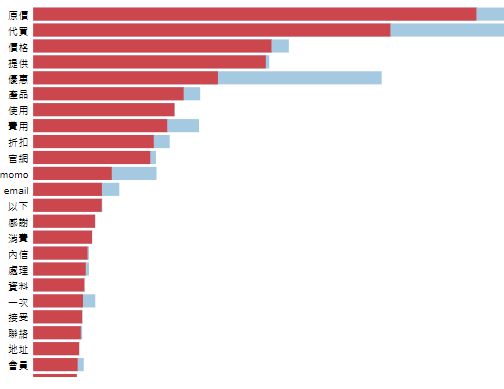
第三類主題:(付款相關)

關鍵字:運費、付款、回饋、台幣



第四類主題:(優惠或價格)

關鍵字:原價、價格、費用、折扣



**五、結論**

本研究結果顯示：(1)當消費者在雙11購物節購物行為的過程中，商品沒有提供優惠、運費需要消費者額外負擔，又或者是自境外購入商品須進行隔離後才能取得，這些行為都會帶來負面情緒。(2)而消費者以遠低於原價的優惠價格取得商品、零售業者以高效率的運送貨品，都會產生較高的正面情緒。由本研究的結果看來，零售業者提供高效率、優惠價格和完美的商品，來誘發消費者購買更有助於消費者產生正面的購後情緒這對零售業者來說是一個好消息。

**六、本組作業心得:**

1.小組初次取得分析資料量約2,000多筆，導致處理時間較冗長，且容易發生作業異常中斷須重新再次執行，後續控制資料量大約1,000筆上下，

再進行後續處理及分析得以讓作業較為順利，也較不會出現錯誤。

2.於WF作業皆以正常執行完後，於視覺化儀表板欲產出圖形化分析，發現在參數設定常出現資料讀取有誤情形，導致無法產出圖形化結果，經多次實驗，須重新建立新的WF流程且最好順利完成後立即進儀表板分析，比較不會發生資料讀取錯誤狀況。