

社群媒體分析 | Group 12

What folks talked about Cafes in Taipei? Utilizing Aspect-based Sentiment Analysis and Eigenvector Centrality Analysis to mine reviews on Google Map

社工四 B09310034 王柏詒 生傳四 B09610020 黃韻文 財金四 B0973047 黃敏瑄

專案分工

組內成員組成多元，不同的背景讓我們可以碰撞出不一樣的觀點和想法



社工四 王柏詒

- 文獻分析
- 資料前處理
- 四大主題分類與高頻詞提取
- **Eigenvector Centrality Analysis**



生傳四 黃韻文

- 文獻分析
- **Google Map Review 資料爬取**
- 資料前處理
- 面向情緒分析模型研究、資料標記、模型訓練與預測



財金四 黃敏瑄

- 四大主題分類研究
- 面向情緒分析結果研究
- 文獻分析
- 簡報製作

研究背景

對於在台北讀書的我們而言，周末下午尋覓一間評價好又具個人風格的咖啡廳讀書是常見的情境，社群平台上亦可見許多人一間一間走訪、辛勤經營的咖啡廳帳號，為的就是讓大眾能按照地區、風格等面向快速的找到自己喜愛的咖啡廳

現況限制

- 評分主觀性
- 評論未分類

預期效益

讓消費者可以
快速找到喜歡的咖啡廳，從結
構化的資訊中掌握各面向的評
價

缺口

目前比較少人針對 Google
map 中文評論做深入的分析

期中提案改善

解決方案

問題

到底是以「文件」
還是「句子」為單位

我們希望可以做 aspect-based analysis，因此以「面向」為分析單位

LDA 產生的 topics
如何對應到 4 大主題

預先建立四大主題字典 (Lexicon)，並查看評論內容是否涵蓋字典詞彙

在四大主題中
找「高詞頻」的目的及作法

想要從行政區和咖啡廳的角度切入，去分析消費者對某間咖啡廳的想法

文獻探討與實作方法

以四大主題
分類與高頻
關鍵字提取

Aspect-
based
Sentiment
Analysis

Eigenvector
Centrality
Analysis

文獻探討

“How was your meal?” Examining customer experience using Google maps reviews. International Journal of Hospitality Management

[Link](#)

A multi-task learning model for Chinese-oriented aspect polarity classification and aspect term extraction

[Link](#)

課堂簡報 Network Essential
Eigenvector centrality tries to generalize degree centrality by incorporating the importance of the neighbors

實作方法

1. 自定義四大字典
2. 根據定義分類主題
3. 從主題看 Frequency

1. 使用文獻作者所提供的開元套件 pyabsa
2. 加入自行標記的資料訓練模型
3. 面向提取並預測情感分數

1. 將評論轉換為矩陣
2. 利用 Eigenvector Centrality 的概念，定義專一度
3. 繪製網路圖

資料來源：爬取 Google Map 評論

1 評論資料在 `listentitle....` 這個連結



2 隨頁面往下滑 render 出來



3 逐一爬取連結直到 review 為空

The screenshot shows the Network tab of a browser's developer tools. The requests are listed under the 'Preview' tab. The first request is a Fetch/XHR for a URL containing 'listentitiesreviews?gl=tw&...'. Subsequent requests are for URLs starting with 'log204?gl=tw&pb=' followed by various parameters. The last two requests are for URLs starting with 'listentitiesreviews?gl=tw&...'. The status bar at the bottom indicates 42 / 124 requests, 151 kB / 389 kB.

Name	Headers	Payload	Preview	Response	Initiator
listentitiesreviews?gl=tw&...					
log204?gl=tw&pb=18m10!...					
log204?gl=tw&pb=18m25!...					
log204?gl=tw&pb=!8m9!3...					
log204?gl=tw&pb=18m10!...					
log204?gl=tw&pb=18m12!...					
log204?gl=tw&pb=18m9!3...					
log204?gl=tw&pb=!8m9!3...					
log204?gl=tw&pb=18m11!...					
log204?gl=tw&pb=18m10!...					
log204?gl=tw&pb=18m9!3...					
listentitiesreviews?gl=tw&...					
listentitiesreviews?gl=tw&...					

資料集說明

行政區	大安、信義、松山、中山、大同、中正
咖啡廳	每區 20 間
資料總筆數	48264

area	store	Name	Time	Review	Star
中山區	1 + 1 Together Republic	"Lu Mjung"	"3 個月前"	"檸檬磅蛋糕好好吃 上面還有有用一點蜂蜜提味 完全會捨不得吃完\n 玫瑰奶茶竟然是真材實料..."	5
中山區	1 + 1 Together Republic	"王思婷"	"4 個月前"	"二訪～一家很有個性、有氣氛的店，店很寬敞，平日下午到店，人不多，為了派和貓咪而來！這裡的派..."	4
中山區	1 + 1 Together Republic	"陳一華"	"1 個月前"	"人多的時候(假日下午/用餐時間)食物需要等比較久，但整體氣氛很好，可以聊天、甚至玩桌遊，很..."	5
中山區	1 + 1 Together Republic	"Katty Chien"	"11 個月前"	"2023/06/05 來消費：\n\n之前在 Google Map 上搜尋貓咪咖啡廳才知道..."	4
中山區	1 + 1 Together Republic	"Vicky77 Chang"	"1 年前"	"咖啡一般般，女店員冷冷的，整家店的標語給人感覺是以客人很煩，不要作妖的前提設立的，但確實也..."	4
...

資料前處理 - 行政區 Frequency、斷詞斷字

① 輸入資料轉換

前處理方式	目的
合併行政區所有咖啡廳資料	以行政區維度分析必須將各個咖啡廳資料集合併共同分析
刪除 Review 空值、非中文評論	使用中文的預訓練模型，不支持英文或日文
根據 \n, 。, ! 切分評論	將同一人的評論透過分割句子瞭解不同句子所描述的面向，而非以「最大值」定義主題

資料前處理

② 輸入資料轉換與訓練結果

前處理方式	目的	總筆數
根據 \n, 。, ! 切分評論	實測句子太長會影響 aspect 萃取的表現，轉換為 sentence-base	157919
去除非中文的評論	使用中文的預訓練模型，不支持英文或日文	124115
使用自訂義的四大面向 aspect-dict 過濾評論	礙於時間限制刪減可能抓不到 aspect 的句子 提取句子類別與關鍵字	79123
刪除少於五個字的句子	高機率只包含 aspect 不包含情緒的句子	69470

以四大主題分類與高頻關鍵字提取

```
# 定義主題詞典
topic_dict = {
    '食品品質': ['咖啡', '咖啡豆', '手沖咖啡', '美式咖啡',
                  '精品咖啡', '甜點', '蛋糕', '起司蛋糕', '回甘',
                  '讚', '好喝', '濃厚', '美味', '口
                  果茶', '蘋果派', '麵包', '微酸', '香',
                  '用料', '餐點', '料理', '好吃'],
    '服務品質': ['態度', '親切', '友善', '熱情', '耐心',
                  '解說', '推薦', '專業', '知識豐富', '熟練',
                  '仔細', '親自', '友好', '講解', '說明',
                  '用心', '等候'],
    '餐廳氛圍': ['明亮', '乾淨', '舒適', '裝潢', '設計',
                  '擁擠', '安靜', '悠閒', '放鬆', '溫馨',
                  '復古', '現代', '翻修', '座位多', '座位少',
                  '氣氛好', '環境優', '裝潢美', '採光', '工],
    '整體價格': ['實惠', '便宜', '貴', '價位', '價錢', '價
                  $', '元', '優惠', '特價', '折扣', '促銷',
                  'cp 值高', '金額', '價位合理', '價錢高']}
}
```

```
def get_topic(sentence):
    if isinstance(sentence, str):
        for topic, keywords in
topic_dict.items():
            if any(keyword in sentence for
keyword in keywords):
                return topic
        return '其他'

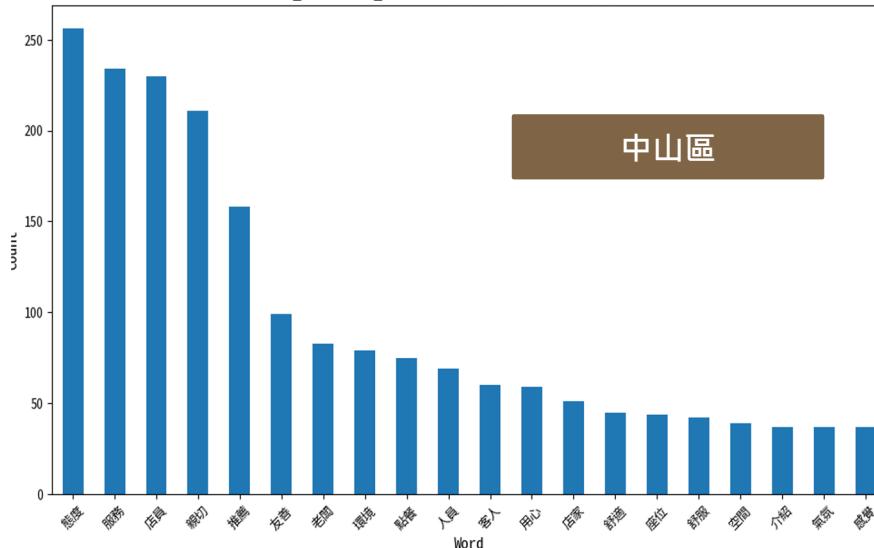
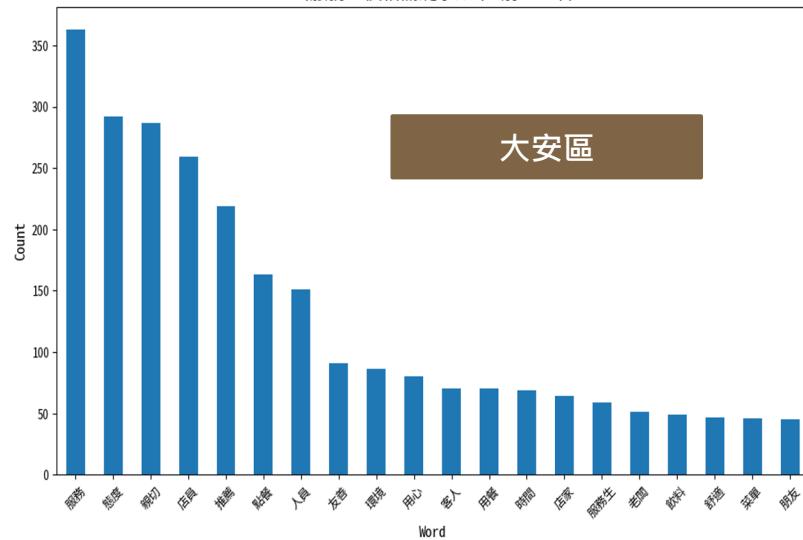
new_df_daan['Topic'] =
new_df_daan['Sentence'].apply(lambda x:
get_topic(x) if isinstance(x, str) else '其他')
```

以四大主題分類 - 大安區為例

'金萱茶 \$180'	整體價格	
'豬肉泡菜蓋飯 \$210 🌶️🌶️'	整體價格	
'BBQ雞肉口袋餅 \$200 🌶️🌶️'	整體價格	
['"人均 : \$700']	整體價格	
		'但很喜歡老屋的氛圍'
		'有插座可充電'
		["台北大安 『2J CAFE』 復古老宅"]
		'當天為假日被安排在入門口第一桌，所以無法拍到2樓的氛圍及空間設計'
		'但主要是整體用餐環境氛圍，以及用餐時間相較於其他餐廳真的蠻不錯的，很適合需要較多時間的客人'
		'👉 餐點份量很足夠！而且都是好吃的'
'☀️ 天氣好的話很適合坐在外面，室內燈光有點昏暗不太適合閱讀～人多的時候會吵點，其他我覺得都...'		
		'餐點也很好吃 整體還是喜歡 ❤️ "]'
		'糖栗子紅茶很香，濃郁好喝，可回沖'
['"18:00以前都不提供訂位，所以中午就早早來等候位'	服務品質	
['"⚠️ 下午五點到很幸運不用等候，但晚上有訂位客人所以有先告知位置只能坐到19:30'	服務品質	
'服務人員也很親切！上菜速度超快的，點完餐20分鐘內都上完 😊'	服務品質	
'店員都很親切'	服務品質	

以四大主題分類與高頻關鍵字提取

- 其實以關鍵字頻率 (frequency) 作為評斷的標準，並不是很嚴謹與具備見解
- 大部分被提到的關鍵字接很相似，這也是我們在本次報告中試圖解決的問題—以情緒分析、面向為切割點作為新的評論指標以取得更多「真實」的評論資訊



Aspect-based Sentiment Analysis : pyabsa 套件



Aspect-based Sentiment Analysis : 模型製作

① 獲取預訓練的中文配置、使用 **FAST_LCF_ATEPC** (基於深度學習) 的模型

```
config = (
    ATEPC.ATEPCCConfigManager.get_atepc_config_chinese()
) # this config contains 'pretrained_bert', it is based on pretrained models
config.model = ATEPC.ATEPCModelList.FAST_LCF_ATEPC
```



② 使用中文資料集，加入自己標記的資料

共抽取 80 則包含不同 aspect 的評論標記，分為測試與訓練集

餐	0	-100
廳	0	-100
氣	B-ASP	Positive
氛	I-ASP	Positive
不	0	-100
錯	0	-100

```
trainer = ATEPC.ATEPCTrainer(
    config=config,
    dataset=my_dataset,
    from_checkpoint="chinese", # if you want to
    auto_device=DeviceTypeOption.AUTO, # use cu
    checkpoint_save_mode=ModelSaveOption.SAVE_M
    load_aug=False, # there are some augmentat
```

Aspect-based Sentiment Analysis : 模型製作

③ 模型訓練結果，根據此訓練模型建立 aspect-extractor

Metric	IQR	Min	Max	Trial	Values	Average	Median	St
Max-APC-Test-Acc w/o Valid Set	2.63	94.74	100	63	[100.0, 94.74]	97.37	97.37	2.
Max-APC-Test-F1 w/o Valid Set	4.38	91.24	100	38	[100.0, 91.24]	95.62	95.62	4.
Max-ATE-Test-F1 w/o Valid Set	15.905	51.52	83.33	905	[83.33, 51.52]	67.425	67.425	15.

- APC (面向極性分類) [90-100] 表現佳且穩定
- ATEPC (面向術語萃取) [52-83] 表現較差，推測為食物面向組合複雜且種類多元

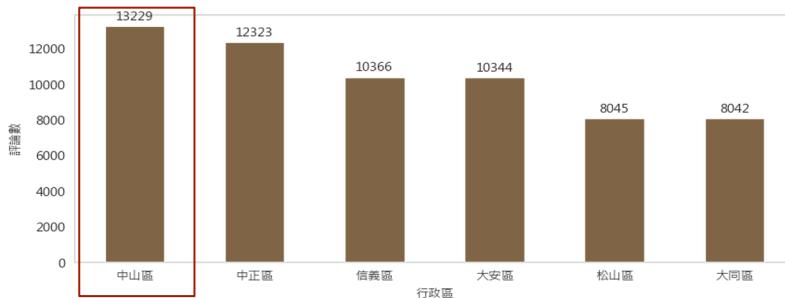
Aspect-based Sentiment Analysis : 預測結果

- 資料筆數從 69470 => 139325
- 面向正規化：統合同個面向的用詞
- 面向過濾：排除掉與四大面向不相關的高頻面向

	area	store	sentence	class	sentiment	confidence	normalized_aspect	new_class
0	中山區	1 + 1 Together Republic	玫瑰奶茶竟然是真材實料的玫瑰花瓣	食物品質	Positive	0.9983	奶茶	食物品質
1	中山區	1 + 1 Together Republic	"二訪～一家很有個性、有氣氛的店，店很寬敞，平日下午到店，人不多，為了派和貓咪而來	食物品質	Positive	0.9986	氣氛	餐廳氣氛
2	中山區	1 + 1 Together Republic	這裡的派很好吃	食物品質	Positive	0.9911	派	食物品質
3	中山區	1 + 1 Together Republic	黑白貓"香香"主動過來，摸貓就開心，店內也有書可以看	餐廳氛圍	Positive	0.9979	貓	餐廳氣氛
4	中山區	1 + 1 Together Republic	黑白貓"香香"主動過來，摸貓就開心，店內也有書可以看	餐廳氛圍	Positive	0.9981	書	餐廳氣氛
5	中山區	1 + 1 Together Republic	老闆娘酷酷的，人蠻好的"	服務品質	Positive	0.9967	老闆娘	服務品質

評論多集中於食物品質，而位於價格和服務意外較少，且中山區評論最多

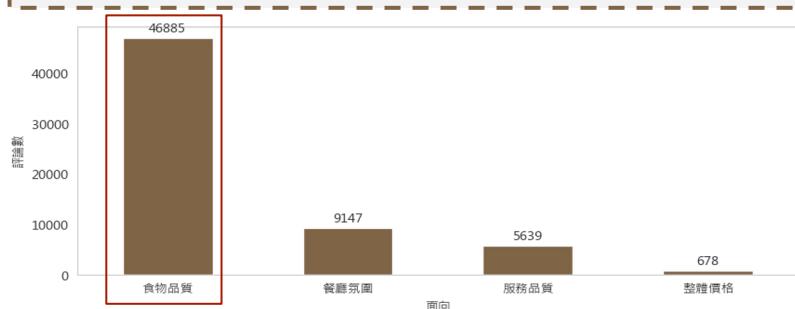
行政區的總評論數



各行政區在四大面向下的評論數

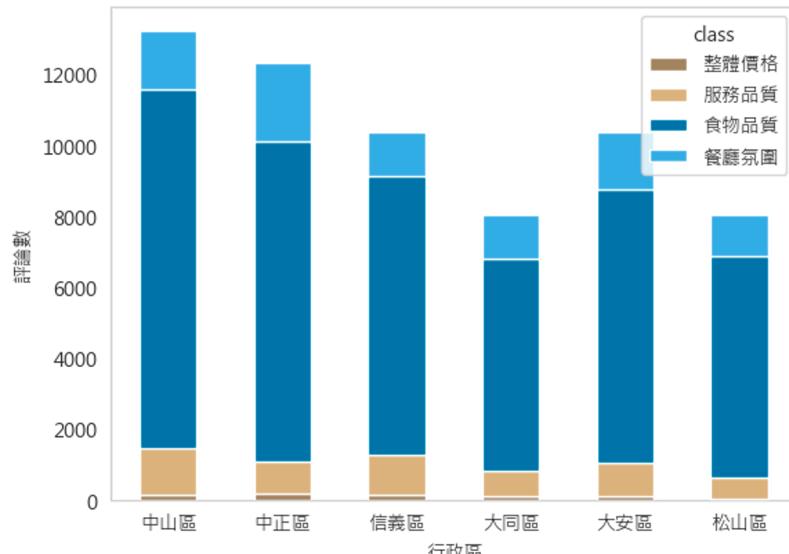
不管在哪一個行政區底下，食物品質的評論都是最多的，反而價格和服務的討論度較低

四大面向的總評論數



class

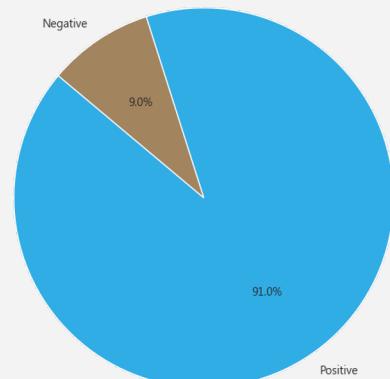
- 整體價格
- 服務品質
- 食物品質
- 餐廳氛圍



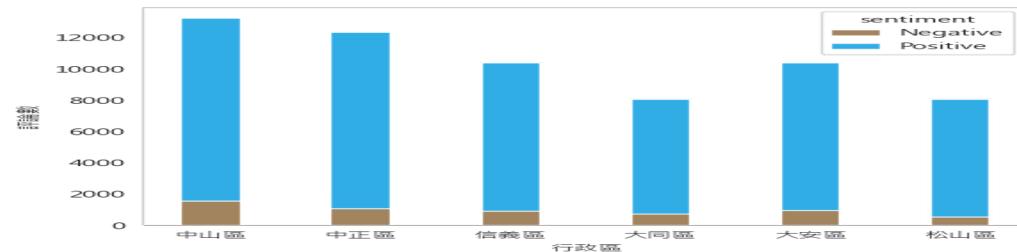
正面評論皆佔大多數，約 90%，情緒分數差距較小，正向回饋的網絡

情緒分數的評論數量分佈

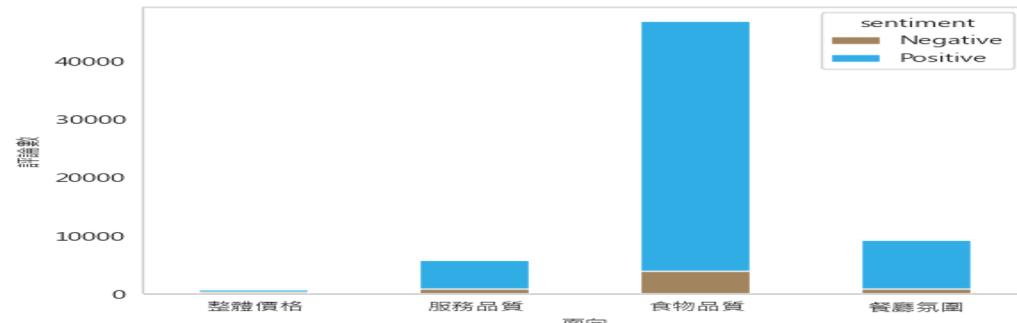
正面評論大約佔了所有評論的 90% 左右，且正負面分數幾乎皆落在 0.9 ~ 1 分之間



各行政區下情緒分數的評論數



四大面向情緒分數的評論數



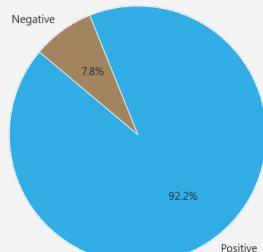
各行政區評論關鍵字在食物品質面向有較大的差異，顯示不同咖啡廳的特色

	中山區	中正區	信義區	大同區	大安區	松山區
食物品質	前二: 蘋果派、奶昔 後二: 果汁、漢堡	前二: 漢堡、冰滴 後二: 深焙、塔	前二: 奶昔、泡芙 後二: 濃縮、卡布	前二: 冰沙、漢堡 後二: 薯條、主食	前二: 摩卡、深焙 後二: 主食、果茶	前二: 餃、濃縮 後二: 主食、糕點
餐廳氛圍	前二: 採光、光線 後二: 動線、座位	前二: 採光、設計 後二: 空間、座位	前二: 光線、採光 後二: 座位、貓	前二: 裝飾、採光 後二: 座位、動線	前二: 動線、採光 後二: 座位、光線	前二: 裝飾、採光 後二: 座位、設計
服務品質	前二: 老闆、店員 後二: 店員、服務	前二: 老闆、服務 後二: 服務、店員	前二: 老闆、店員 後二: 店員、服務	前二: 老闆、店員 後二: 店員、服務	前二: 老闆、服務 後二: 服務、店員	前二: 老闆、服務 後二: 服務、店員
整體價格	前二: 費用、性價比 後二: 金額、單價	前二: 金額、性價比 後二: 價格、單價	前二: 費用、性價比 後二: 消費、金額	前二: 單價、價格 後二: 性價比、消費	前二: 金額、單價 後二: 性價比、費用	前二: 消費、價格 後二: 單價、費用

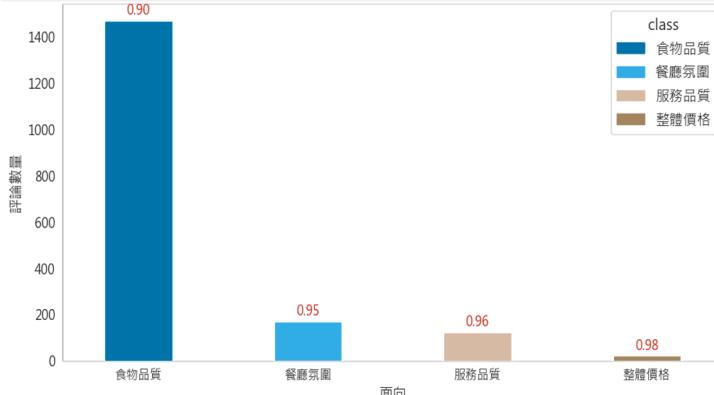
以「興波咖啡」為例進行詳細說明

正負面評論數量分佈

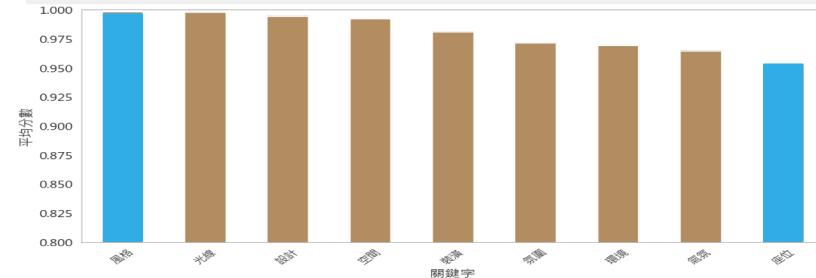
正面評論大約佔了所有評論的 **90%** 左右，且正負面分數幾乎皆落在 0.9 ~ 1 分之間



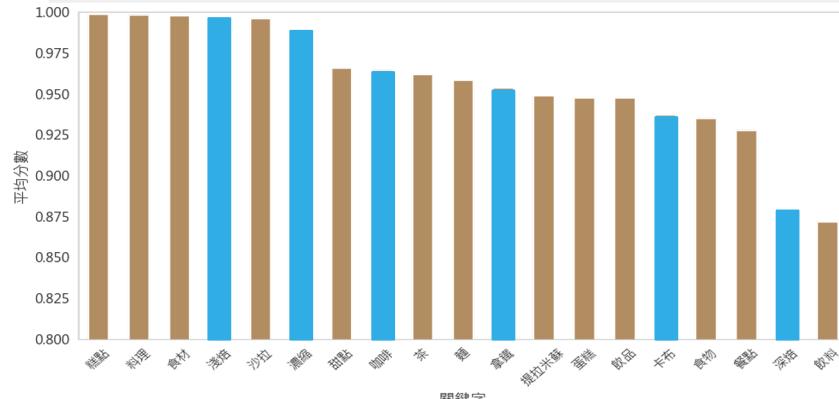
四大面向評論數量與其相應情緒分數



餐廳氛圍下的關鍵字



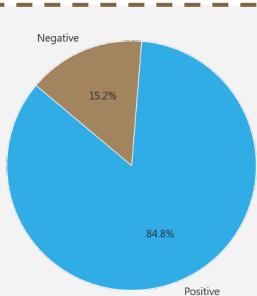
食物品質下的關鍵字



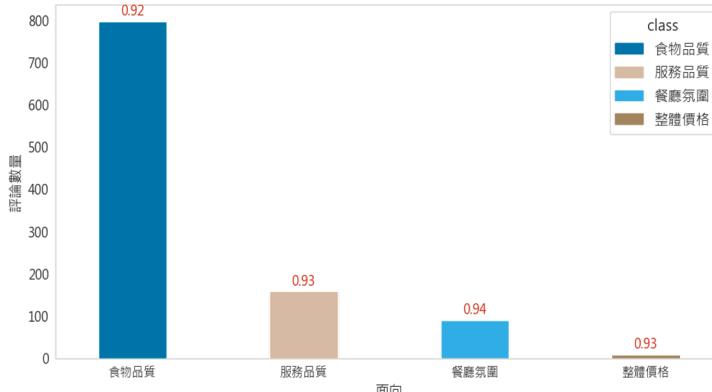
以「佐曼咖啡館」為例進行詳細說明

正負面評論數量分佈

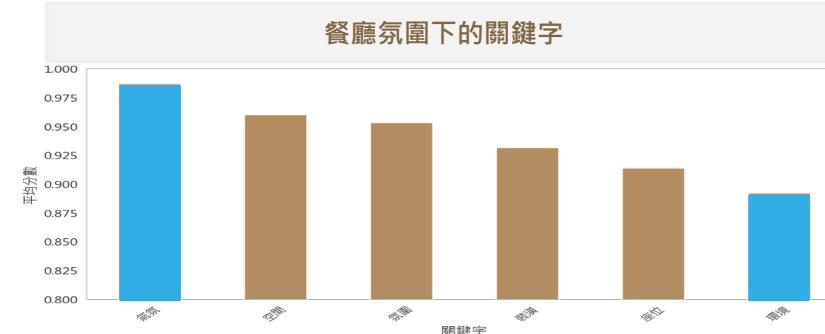
正面評論大約佔了所有評論的 **85%** 左右，且正面分數幾乎皆落在 0.9 ~ 1 分之間



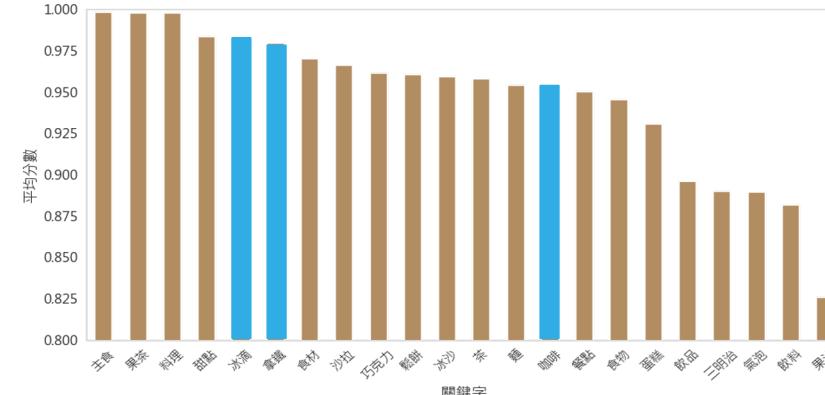
四大面向評論數量與其相應情緒分數



餐廳氛圍下的關鍵字



食物品質下的關鍵字



Eigenvector Centrality Analysis：哪些在消費者中存在高度專一性 / 忠誠度

為了找出消費者與咖啡廳的互動網絡，並瞭解是否在 Google Reviews 上也存在關鍵消費者（KOC），我們需要將評論轉換為矩陣。定義如下：



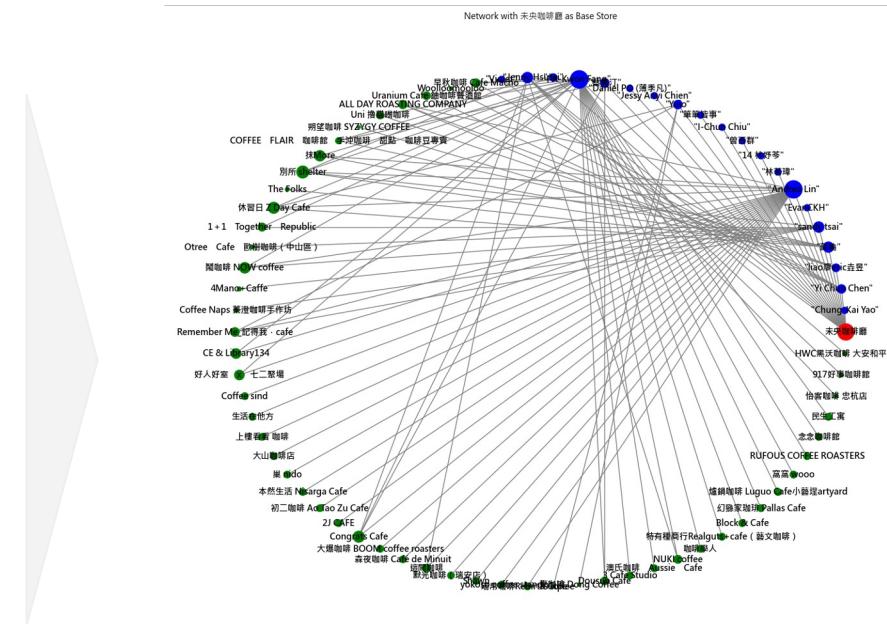
Eigenvector Centrality Analysis : Eigenvector Centrality 與評論者專一度

咖啡店	Eigenvector Centrality	評論者	專一度
1 + 1 Together Republic	0.04638048455276768	Lu Mjung	0.2
3 Cafe Studio	0.06973690666750512	王思婷	1.0
百花台	0.16903020953841447	陳妍儒	0.5
生活在他方	0.14952612626683426	Vincent Li	0.16666666666666666
民生工寓	0.13436142770409973	c2v2v2	0.08333333333333333

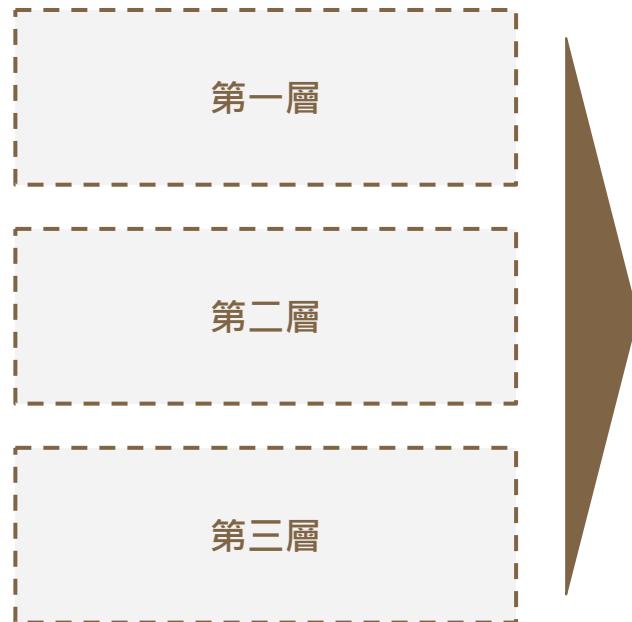
Eigenvector Centrality Analysis : 由網路圖瞭解評論者與咖啡廳之間的關係

以未央咖啡廳為例：

1. 第一層以某一咖啡廳為中心，向外連結
 1. 第二層隨機選擇 20 位評論過該咖啡廳的評論者
 1. 第三層為這 20 位隨機評論者曾經還評論過的其他咖啡廳



Eigenvector Centrality Analysis：由網路圖瞭解評論者與咖啡廳之間的關係



瞭解該咖啡廳的消費者輪廓是為「專一度高者」還是「跑咖性質者」 -> 以經營者的角度來說，消費者專一度不高的咖啡廳，不需要時常更換菜單以維繫與舊客的新鮮感即能不間斷地吸引消費者

部分喜愛跑咖的消費者，我們可以將之定義為 **Google Reviews** 上的咖啡廳關鍵消費者（KOC），未來若以投放廣告、素人宣傳方向思考，可優先連結

從關鍵消費者的評論路徑中，找到類似於關連規則的咖啡廳

未來建議

標記更多資料

釐清無法抓取的面向背後的邏輯
，使用更準確的資料標記

- ATEPC 面向術語萃取
- APC 面向極性分類

匯入網站

以 **python** 為後端製作台北市咖啡廳主題地圖，提升專案實際應用性，並能根據使用者回饋調整資料分析的面向，呈現方式等資訊

爬取更多咖啡廳資訊

爬取更多間咖啡廳的資訊，讓 Eigenvalue Centrality 的分析更準確公正。也可以嘗試優化網絡的計算方式，目前考量運算資源與視覺化，未將所有曾經與咖啡廳互動的評論者網絡繪製出來，未來可思考如何優化網絡算法。

Q & A

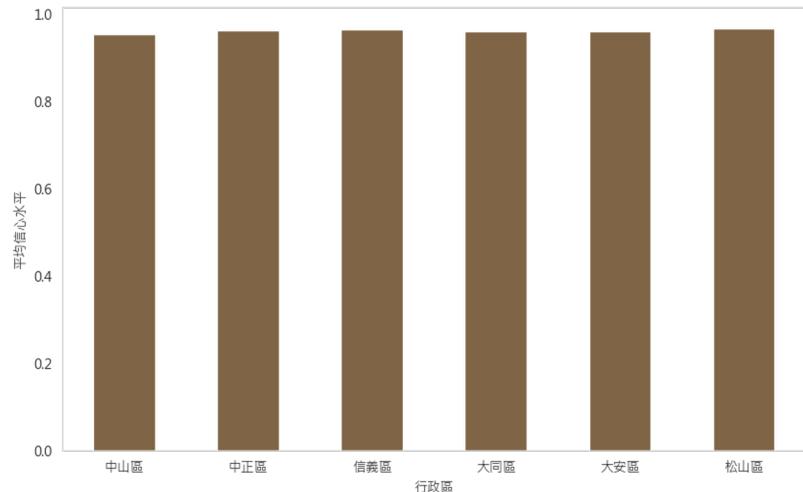
附錄 - Aspect-based Sentiment Analysis : 預測資料前處理

自定義主題辭點

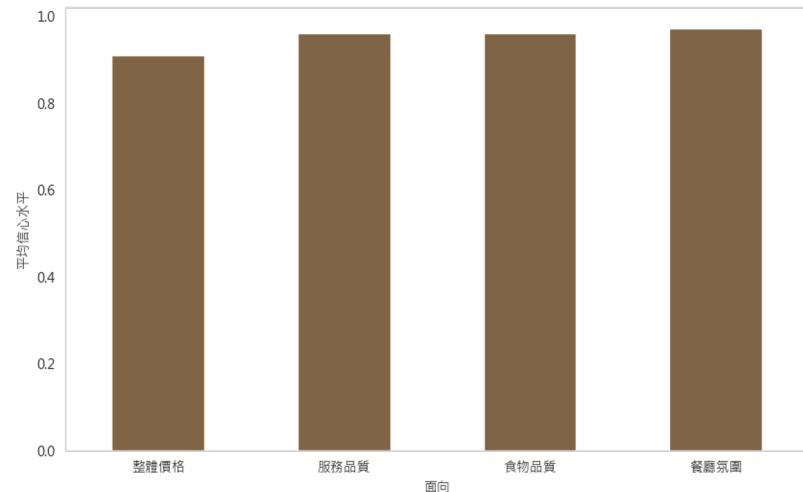
```
# 定義主題詞典
topic_dict = {
    '食品品質': ['咖啡', '拿鐵', '深焙', '淺焙', '甜點', '蛋糕', '冰滴',
                  '果茶', '蘋果派', '食材', '飲料', '飲品', '糕點', '主食', '早午餐',
                  '用料', '餐點', '料理', '提拉米蘇', '可頌', '泡芙', '布丁', '布朗尼',
                  '派', '塔', '蘋果派', '馬卡龍', '卡布', '美式', '摩卡', '濃縮', '冰沙', '麵', '餃', '餐點', '食物', '茶', '沙拉',
                  '抹茶', '巧克力', '奶昔', '果汁', '氣泡', '三明治', '鬆餅', '漢堡', '薯條', '水果杯', '蛋糕'],
    '服務品質': ['服務', '店員', '老闆'],
    '餐廳氛圍': ['裝潢', '設計', '風格', '氛圍', '動線', '空間', '座位', '光線', '裝飾', '座位',
                  '氣氛', '環境', '採光', '貓'],
    '整體價格': ['性價比', '金額', '單價', '總價', '價格', '消費', '費用']}
}
```

附錄 - Aspect-based Sentiment Analysis : 評分信心水平分佈集中在 0.9~1

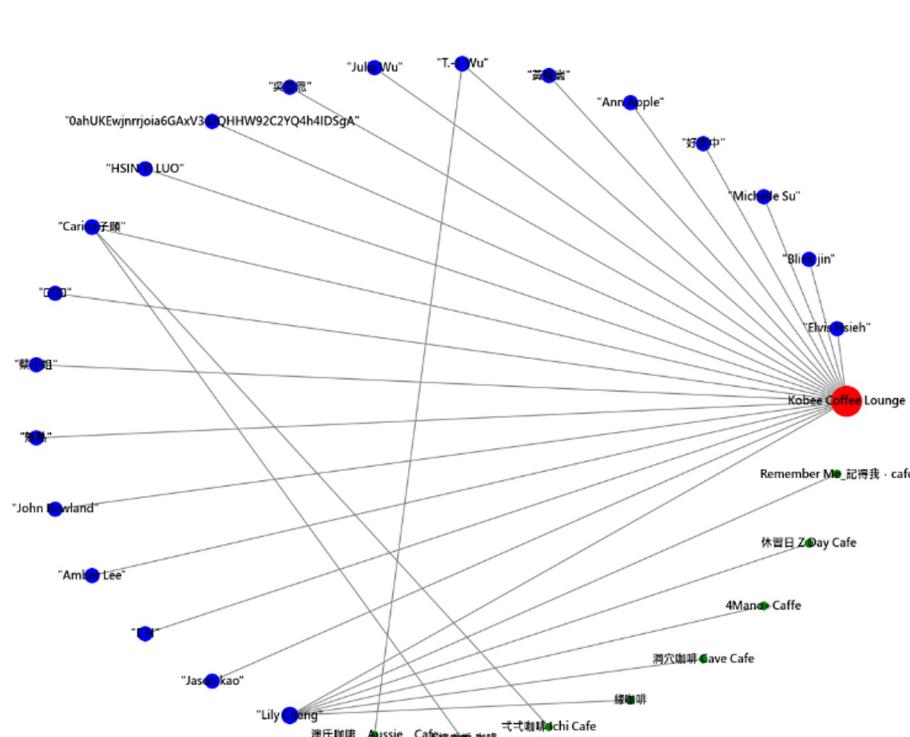
不同行政區的平均信心水平



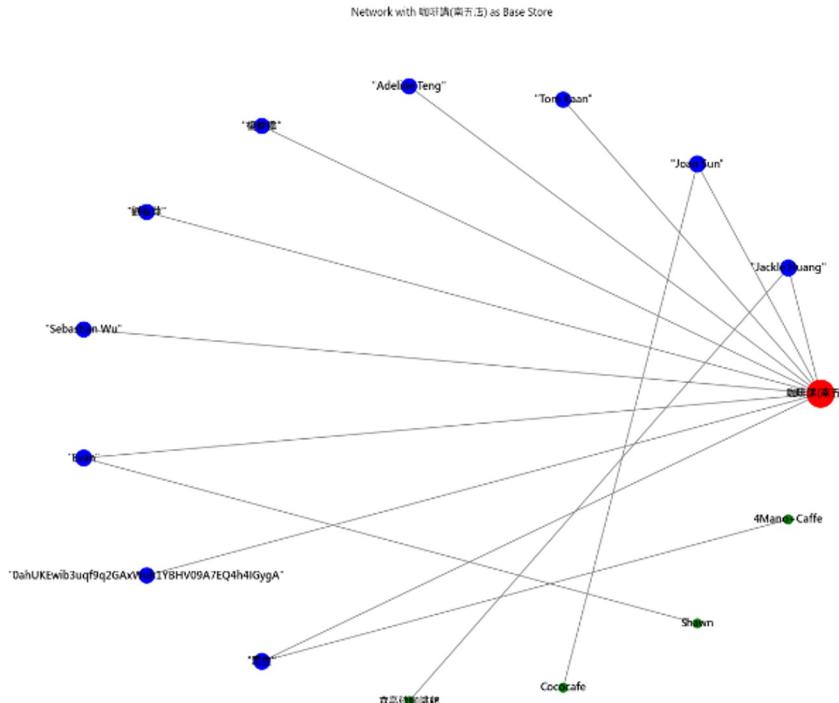
不同面向的平均信心水平



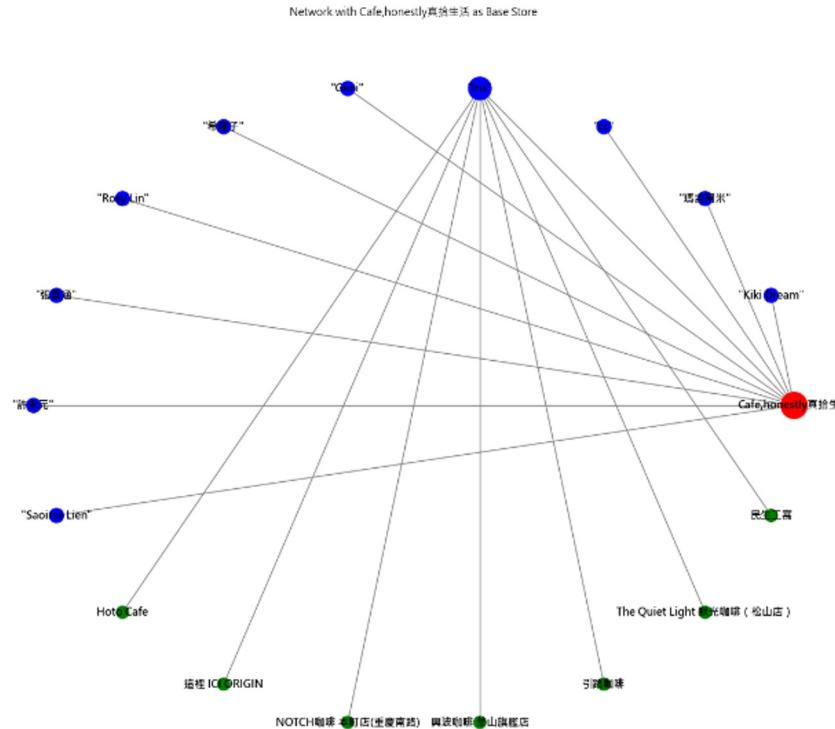
附錄 - Eigenvector Centrality Analysis : Kobee Coffee Lounge



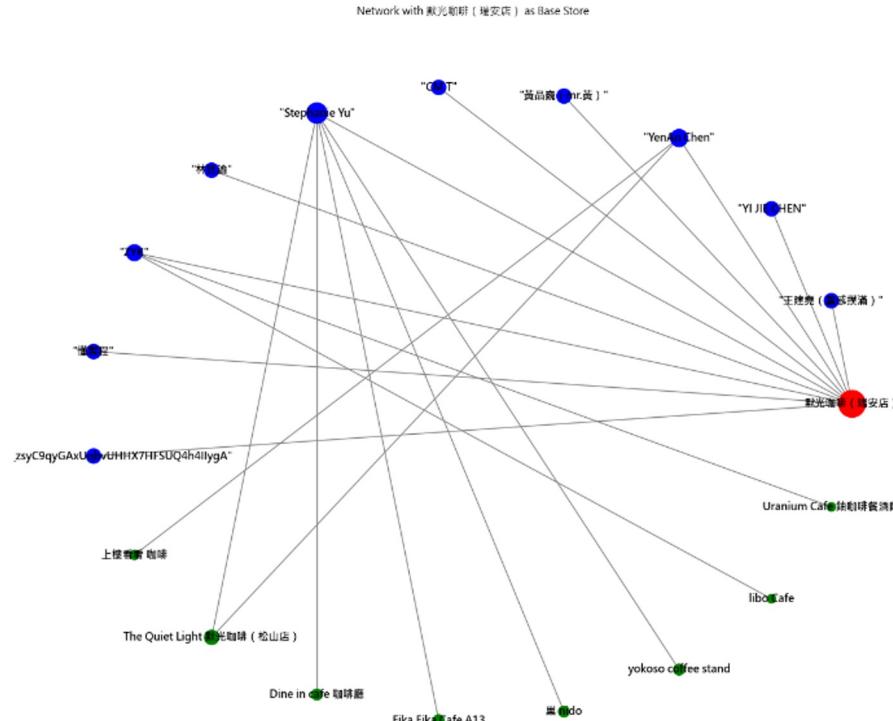
附錄 - Eigenvector Centrality Analysis : 咖啡講 (南五店)



附錄 - Eigenvector Centrality Analysis : Honestly 真拾咖啡

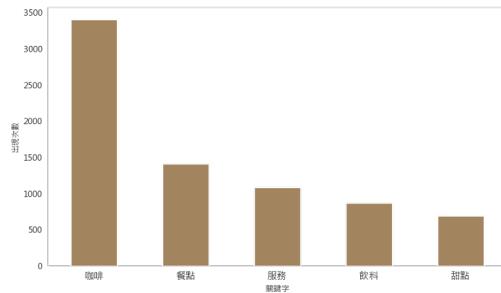


附錄 - Eigenvector Centrality Analysis : 默光咖啡 (瑞安店)

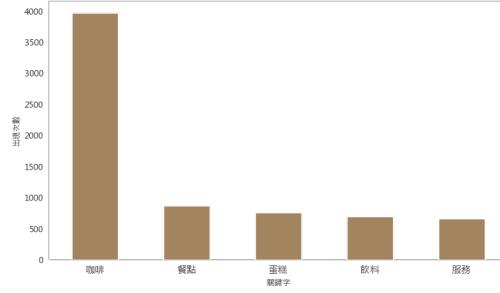


各行政區評論關鍵字集中於咖啡、餐點和服務

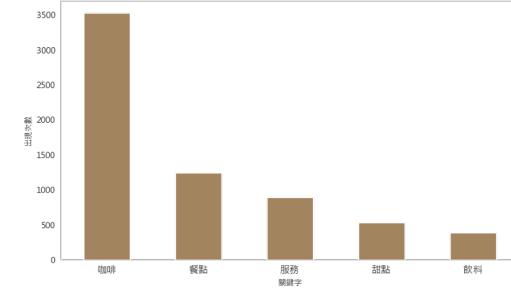
中山區：咖啡、餐點、服務



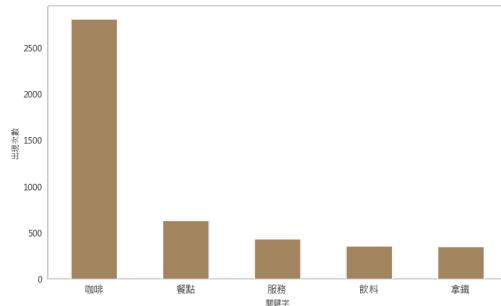
中正區：咖啡、餐點、服務



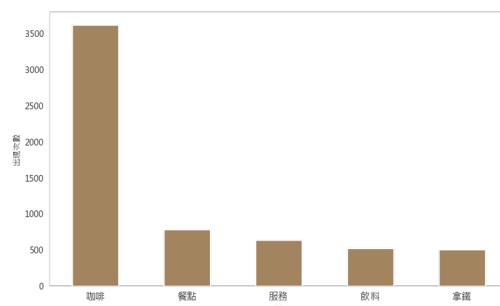
信義區：咖啡、餐點、服務



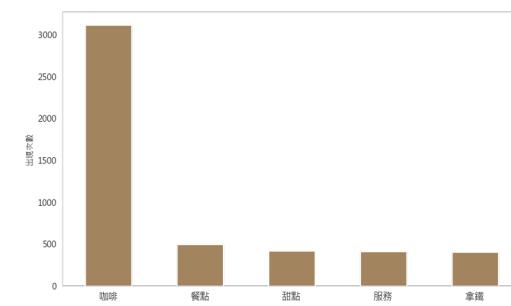
大同區：咖啡、餐點、服務



大安區：咖啡、餐點、服務

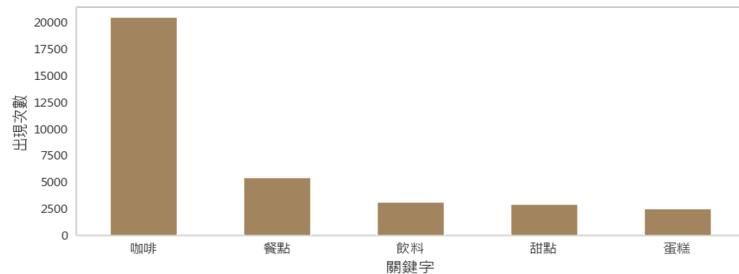


松山區：咖啡、餐點、甜點

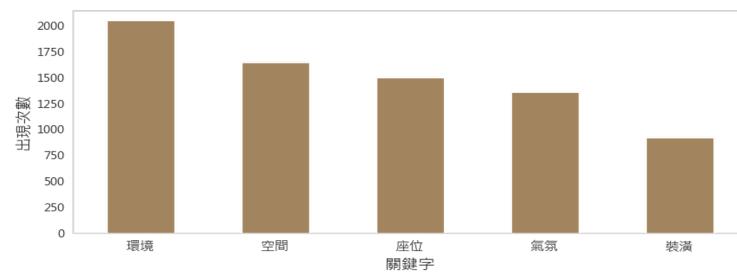


從四大面向內最常見的關鍵字可以看出主題分類準確

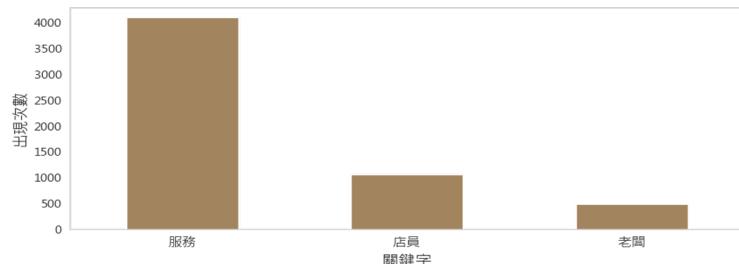
食物品質：咖啡、餐點



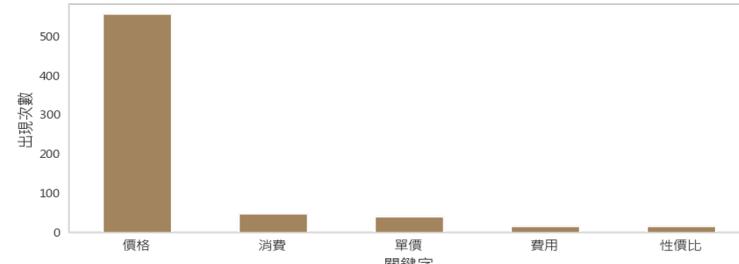
餐廳氛圍：環境、空間



服務品質：服務、店員



整體價格：價格、消費



Eigenvector Centrality Analysis : 哪些在消費者中存在高度專一性 / 忠誠度

將評論轉換為矩陣

	1 + 1 Together Republic	3 Cafe Studio	4MANO CAFFÉ X classic	CE & Library134	COFFEE FLAIR 咖啡館 手沖咖啡 甜點 咖啡豆專賣	Dine in cafe 咖啡廳	Hoto Cafe	Ibo Cafe	Otree Cafe 歐樹咖啡 (中 山區)	佐 曼 咖啡 館	...	yokoso coffee stand	他。植 物咖啡 He Coffee	咖啡 講 (南 五 店)	四維咖啡 商號 Cafe4way	念 念 咖啡 館	手沖 咖啡 植咖 啡 自家烘 焙 Drip Coffe コ一 ヒ一 Café	暫停工 作室 PAUSE Studio	民 生 工 寓	河 光 咖 啡 屋	瑞 Re Co
"Lu Mjung"	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	
"王思婷"	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	
"陳一華"	1	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	0	0	0	
"Katty Chien"	1	0	0	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	
"Vicky77 Chang"	1	0	0	0	0	0	0	0	1	0	0	...	0	0	0	0	0	0	0	0	

References

- Mathayomchan, Boonyanit & Taecharungroj, Viriya. (2020). "How was your meal?" Examining customer experience using Google Maps reviews. International Journal of Hospitality Management.
- Heng Yang, Binqing Zeng, JianHao Yang, Youwei Song, Ruyang Xu (2019) A Multi-task Learning Model for Chinese-oriented Aspect Polarity Classification and Aspect Term Extraction
- Chien Chin Chen (2024) Network Essentials: Eigenvector Centrality