

餐飲分析

CAMA V.S. LOUISA

Sophia, Blaire, Joelle, Amy

目標客戶



本報告旨在分析 CAMA與LOUISA
兩大本土咖啡品牌：

1. 透過消費者回饋與心得, 進行深入的資料收集與分析
2. 為目標客戶提供有價值的市場洞察

資料蒐集方向

爬蟲方
法

資料存
儲

清理與分
析

視覺化

1. 專案目標定位及分析方向

>> 客戶關心什麼？

- 品牌聲量與競爭力
- 消費者口碑
- 市場機會

>> 選擇資料：

- 專業數位媒體
- 社群平台
- 影音媒體
- 大量網路評論資料

資料蒐集方向

爬蟲方
法

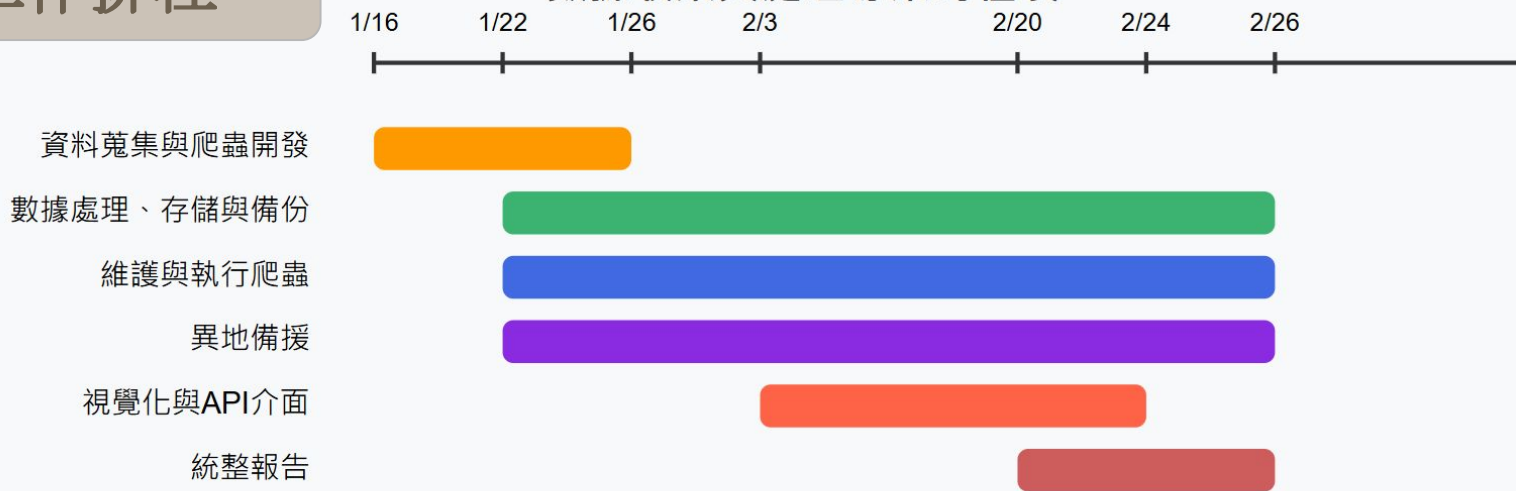
資料存
儲

清理與分
析

視覺化

2. 工作排程

數據收集與處理專案時程表



圖例

- 資料蒐集與爬蟲開發 (1/16 - 1/26)
- 數據處理、存儲與備份 (1/22 - 2/26)
- 維護與執行爬蟲 (1/22 - 2/26)
- 異地備援 (1/22 - 2/26)
- 視覺化與API介面 (2/3 - 2/24)
- 統整報告 (2/20 - 2/26)

3. 任務分工

負責人	爬蟲程式、資料清洗	主要任務
Sophia	Google評論	<ul style="list-style-type: none">- 爬蟲主機排程- API swagger 介面架設- 共用 GitHub 管理
Blaire	Dcard PTT	<ul style="list-style-type: none">- 資料庫主機管理- 製作視覺化分析－儀表板- 撰寫專案行政文件
Joelle	YouTube	<ul style="list-style-type: none">- 會議記錄- 製作視覺化分析－圖表- 簡報統籌、發表
Amy	FoodNext網路媒體	<ul style="list-style-type: none">- 架設GCP雲端異地備援- 簡報初稿編輯- 製作視覺化分析－圖表

介紹

- 深入分析報導的專業媒體
- 業界人士、品牌行銷人員



- 年輕消費者
- 互動活躍
- 較多品牌體驗分享

- 台灣最大匿名論壇
- 更深度、長篇的社群討論



- 提供地理數據
- 顧客體驗的星級評價



Google Maps



- 全球最大影音平台
- 大量討論熱度和曝光度

資料蒐集方向

爬蟲方法

資料存儲

清理與分析

視覺化

食力 FoodNext

- Selenium 自動瀏覽網站
- 以關鍵字搜尋並擷取文章 內容

The screenshot shows the FoodNext website interface. At the top, there is a navigation bar with the site logo '食力 foodNEXT' and links for '食專題', '食科學', '食新聞', and '食生活'. Below the navigation bar, a search bar contains the keyword 'cama'. The search results section displays the title of an article: '【cama】查詢結果，共127筆'. Below the title, there is a list of search results, with the first one being an article titled '【懶人包】9天春節連假難以收心？開工補班優惠大放送幫你擺脫收假症候群！'. The date '2025/02/06' and the URL 'https://www.foodnext.net/life/recipes/lunch/paper/6971053931' are also visible.

食力 foodNEXT 食生活

食專題 食科學 食新聞

HOME / 查詢列表

cama

【cama】查詢結果，共127筆

1. 文章 【懶人包】9天春節連假難以收心？開工補班優惠大放送幫你擺脫收假症候群！

© 2025/02/06 <https://www.foodnext.net/life/recipes/lunch/paper/6971053931>

結束春節連假，要從假期的慵懶回到忙碌的生活軌道，開工的第一步總顯得有些沉重。各大超商與餐館紛紛推出開工優惠，為每一位

資料蒐集方向

爬蟲方法

資料

Google Maps

- Selenium 模擬點擊評論
- 兩品牌各抽樣 50家門市
- 選擇最新排序 內容

```
for review in sel
    review_id = r
    if review_id
        continue

    try:
        try:
            conte
        except:
            conte

    rating =
        i

    brand = "
    content_t
    crawling_
```

三 cama café 台東鐵花村店

評論

全部 拿鐵 9 香 8 店員 5 +6

最相關 最新 評分最高 評分最低

Amily
在地嚮導 · 98 則評論 · 2,468 張相片

★★★★☆ 4 個月前

去年開門只有咖啡
今年有麵包
高雄蜂巢麵包
皮脆內紮實
真的不錯
檸檬冷萃也好喝
就是內用還是用紙杯
少了泡咖啡廳的感覺
紙杯當外帶杯是不錯
下次還要嚐嚐其他的麵包。

資料蒐集方向

爬蟲方法

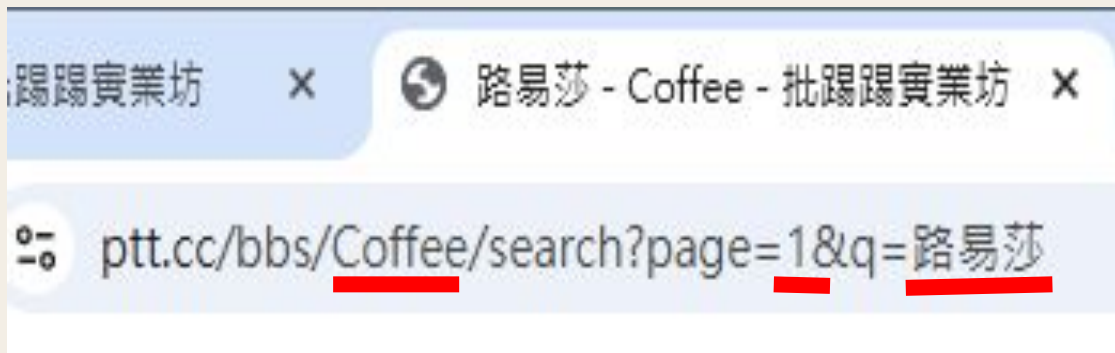
資料存儲

清理與分析

視覺化

PTT

- 使用 BeautifulSoup
- 根據網址規律變更搜尋參數



看板名稱

頁數

關鍵字

```
# 定義咖啡店關鍵字
self.cafe_keywords = {
    "cama": [
        "cama",
        "CAMA",
        "Cama",
        "CAMA咖啡",
        "CAMA café",
        "CAMA coffee",
    ],
    "路易莎": [
        "路易莎",
        "louisa",
        "LOUISA",
        "Louisa",
        "LOUISA COFFEE",
        "Louisa Coffee",
        "路易莎咖啡",
        "路易莎門市",
    ],
}
```

資料蒐集方向

爬蟲方法

資料存儲

清理與分析

視覺化

Dcard

- 取得 Google Custom Search API Key及自訂化引擎 ID
- 爬取Google搜尋引擎介面對關鍵字的摘要
- 不直接進到 Dcard 網站中

CAMA 關鍵字

約有 4,260 項結果 (搜尋時間：0.39 秒)

搜尋摘要

cama | Dcard
Dcard > topics > cama

cama. 驗車完之後來到cama喝杯飲料。cama店內的擺設，點了蜂蜜風味奶茶，我覺得cama這個蠻好喝的，不會很甜同時喝到蜂蜜、奶茶的香味。另一杯是輕焙烏龍。

請問桃園區的**CAMA** 上半年怎麼收了三家？ - 看板Taoyuan
Ptt 批踢踢實業坊 > bbs > Taoyuan > M.1684755187.A.CF7.html

2023年5月22日 ... 年初的時候、靠近縣政府的中山店收店了再來是大興店今天在Uber eat上發現怎麼沒跳出在光南一樓的中華店、一查才發現也停業了怎麼不到半年的時間、有 ...

[其它] **cama**的量販豆品質比意料中的好- 看板Coffee
Ptt 批踢踢實業坊 > bbs > Coffee > M.1649491080.A.D72.html

2022年4月9日 ... 我個人習慣不吃早餐只手沖一杯黑咖啡當一天的開始(所以品質差的咖啡會害我胃食道逆流) 風味比較偏好中~中深焙堅果調性的風味偶爾穿插一下衣索比亞的淺 ...

資料蒐集方向

爬蟲方法

資料存儲

清理與分析

視覺化

YouTube

- 使用 YouTube Data API v3 API Key
- 支援 API 金鑰輪換制
- 以關鍵字搜尋影片並抓取留言

```
def retry_request(request_function, retries=3, delay=5):  
    """帶有重試機制的 API 請求"""  
    for attempt in range(retries):  
        try:  
            return request_function()  
        except HttpError as e:  
            if "quotaExceeded" in str(e):  
                print("API 配額已用完，切換 API Key 後重試...")  
                os.environ['YOUTUBE_API_KEY'] = get_next_api_key()  
            else:  
                print(f"發生錯誤: {e}")  
            if attempt < retries - 1:  
                print(f"重試第 {attempt + 1} 次...")  
                time.sleep(delay)  
            else:  
                raise
```

資料蒐集方向

爬蟲方
法

資料存
儲

清理與分
析

視覺化

1. 爬蟲機

- 配置相關網站資料夾
- 定期執行爬蟲程式
- 程式碼皆設置避免重複儲存相同 內容
- 定期備份爬蟲機 內容(backup)

```
$ cd backups  
$ ls  
CLEAN_SOPHIA20250220.sql.gz  
CLEAN_SOPHIA20250225.sql.gz  
SOPHIA20250205.sql.gz  
SOPHIA20250212.sql.gz  
SOPHIA20250220.sql.gz  
SOPHIA20250225.sql.gz
```

```
$ ls  
1      backups  foodnext  my_env  youtube  
API    dcard     google_map ptt
```

資料蒐集方向

爬蟲方
法

資料存
儲

清理與分
析

視覺化

2. 地端資料庫：MariaDB

- 建立兩個資料庫，一個連接爬蟲機，另個放置經清洗處理的資料

- 定期備份資料庫 內容至雲端

```
MariaDB [(none)]> show databases;
```

Database
CLEAN_SOPHIA
SOPHIA
information_schema

```
MariaDB [SOPHIA]> show tables;
```

Tables_in_SOPHIA
dcard
dcard_2
foodnext_cama
foodnext_louisa
google_map
ptt
ptt_2
ptt_crawled_pages
youtube_cama
youtube_louisa

```
MariaDB [CLEAN_SOPHIA]> show tables;
```

Tables_in_CLEAN_SOPHIA
dcard
foodnext_article_cama
foodnext_article_louisa
google_map
google_map_rating_count
google_map_statistics
google_map_store_region
ptt
youtube_comments_cama
youtube_comments_louisa

資料蒐集方向

爬蟲方
法

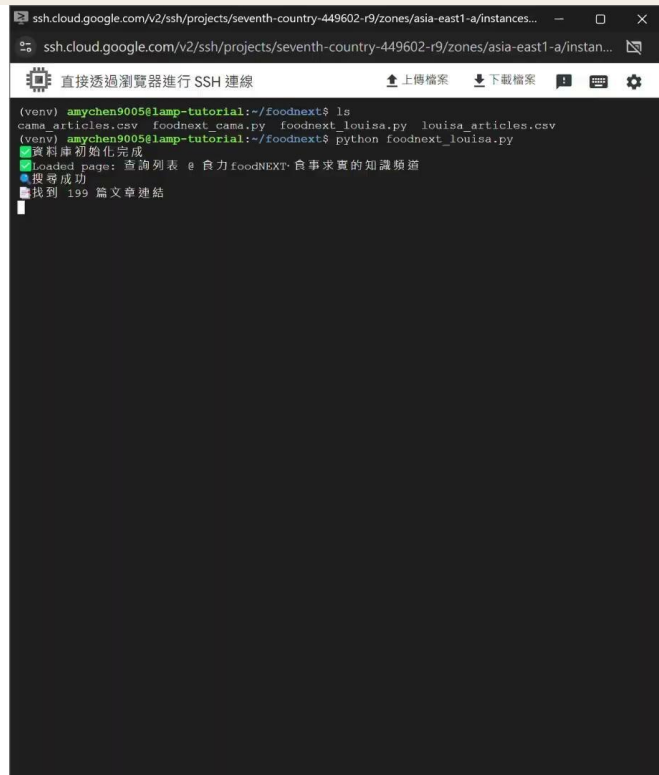
資料存
儲

清理與分
析

視覺化

3. 異地備援：GCP

- 在雲端環境建構一條龍環境
- 排程安排在爬蟲機備份後執行
- DEMO 運行爬蟲測試



The screenshot shows a terminal window connected to a GCP instance. The user runs a series of commands to test a web crawler. The output shows that the crawler successfully initialized, loaded a page, and found 199 article links.

```
ssh.cloud.google.com/v2/ssh/projects/seventh-country-449602-r9/zones/asia-east1-a/instances...  
ssh.cloud.google.com/v2/ssh/projects/seventh-country-449602-r9/zones/asia-east1-a/instan...  
直接透過瀏覽器進行 SSH 連線  
上機檔案 下載檔案  
(venv) amychen9005@lamp-tutorial:~/foodnext$ ls  
cama_articles.csv foodnext cama.py foodnext_louisa.py louisa_articles.csv  
(venv) amychen9005@lamp-tutorial:~/foodnext$ python foodnext_louisa.py  
資料庫初始化完成  
Loaded page: 查詢列表 @ 食力foodNEXT·食事求實的知識頻道  
搜尋成功  
找到 199 篇文章連結
```

資料蒐集方向

爬蟲方
法

資料存
儲

清理與分
析

視覺化

清洗程 式

- 結構化資料 內容 (統一欄位)
- 針對使用者評論的平台做語意分析的資料處理

資料蒐集方向

爬蟲方
法

資料存
儲

清理與分
析

視覺化

1. PowerBI 視覺化分析圖表

運用以下方式實時連接遠端 MariaDB :

- 安裝ODBC驅動程式
- 配置 DSN
- 開SSH隧道



資料蒐集方向

爬蟲方
法

資料存
儲

清理與分
析

視覺化

2. API 架設

利用 API 直接取得資料庫資料

以品牌名稱和地區作為 查詢的選擇

<http://dv107.coded2.fun:10006/swagger-ui/>

Scan me!



API swagger website

Database data show ^{v1}

/swagger/

Google_map



GET /GoogleMapRatingCount

GET /GoogleMapStatistics

GET /GoogleMapStoreRegion

GET /google_map

Dcard



GET /dcard

Foodnext



GET /foodnext_cama

GET /foodnext_louisa

Ptt



GET /ptt

Youtube



GET /youtube_cama

GET /youtube_louisa

API swagger website

GET /google_map	
獲取 Google Map 評論資料，可依品牌、地區和評分進行篩選	
Parameters 以 Google Maps 為例	
Name	Description
brand * required string (query)	品牌名稱 Available values : Cama, Louisa
region string (query)	地區 Available values : 北部, 中部, 南部, 東部, 外島
rating string (query)	評分 Available values : 1, 2, 3, 4, 5

API swagger website

GET /GoogleMapRatingCount

獲取 Google Map 評分統計資料，可依品牌、地區和評分進行篩選

Parameters

Name	Description
brand * required string (query)	品牌名稱 <div>Cama</div>
region string (query)	<div>Cama</div> <div>Louisa</div>
rating string (query)	北部 評分 <div>5</div>

API swagger website

Code

Details

200

Response body

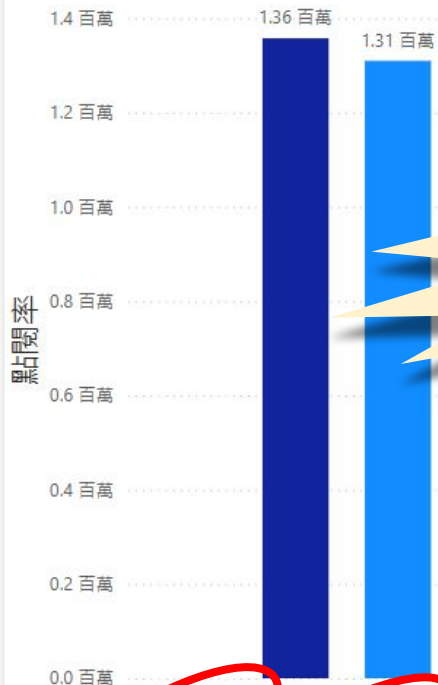
```
{
  "data": [
    {
      "brand": "Cama",
      "content": "cama店員訓練很好，廁所保持的很乾淨",
      "content_time": "4 週前",
      "id": 8265,
      "rating": 5,
      "region": "北部",
      "store_name": "羅東公正店"
    },
    {
      "brand": "Cama",
      "content": "色調溫暖的咖啡廳，咖啡好喝、服務簡單親切、音樂輕鬆，極好放空的地方",
      "content_time": "1 個月前",
      "id": 8267,
      "rating": 5,
      "region": "北部",
      "store_name": "羅東公正店"
    },
    {
      "brand": "Cama",
      "content": "店內的布置很溫馨，是讀書、休憩的好地方，可惜我愛的肉桂捲正好沒有了。之前看過他們的菜單，三明治看起來也挺不錯的。今天因為定座坐包廂，很棒的空間，而且離羅東車站不遠，希望下次再來。? _",
      "content_time": "5 個月前",
      "id": 8273,
      "rating": 5,
      "region": "北部",
      "store_name": "羅東公正店"
    }
  ]
}
```

顯示資料庫的內容：
品牌、留言、留言時間、id、評分、地區、分店名

Download

Response headers

```
connection: close
content-length: 582961
content-type: application/json
date: Wed, 26 Feb 2025 03:23:53 GMT
server: Werkzeug/2.2.2 Python/3.13.1
```



食安、品質！??

西雅圖摻混是被騙還是蓄意？拆解台灣咖啡豆供應鏈看仔細

品牌老化、畜牧污染、包裝陳舊等六大困境纏身！飲食產業如何升級再造？

西雅圖摻混是被騙還...
品牌老化、畜牧污染、包裝...
4種咖啡職人、2種證...
黑金

平台數

5

資料總數

13,077

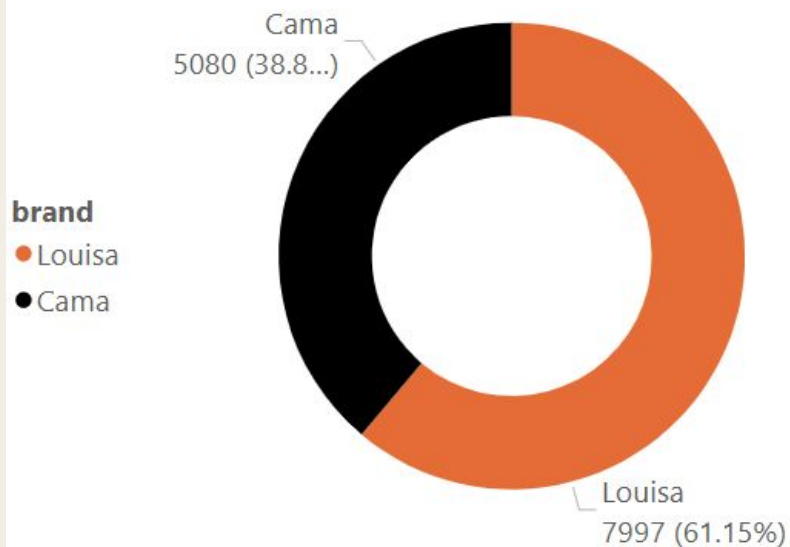
最早資料

2021年1月6日

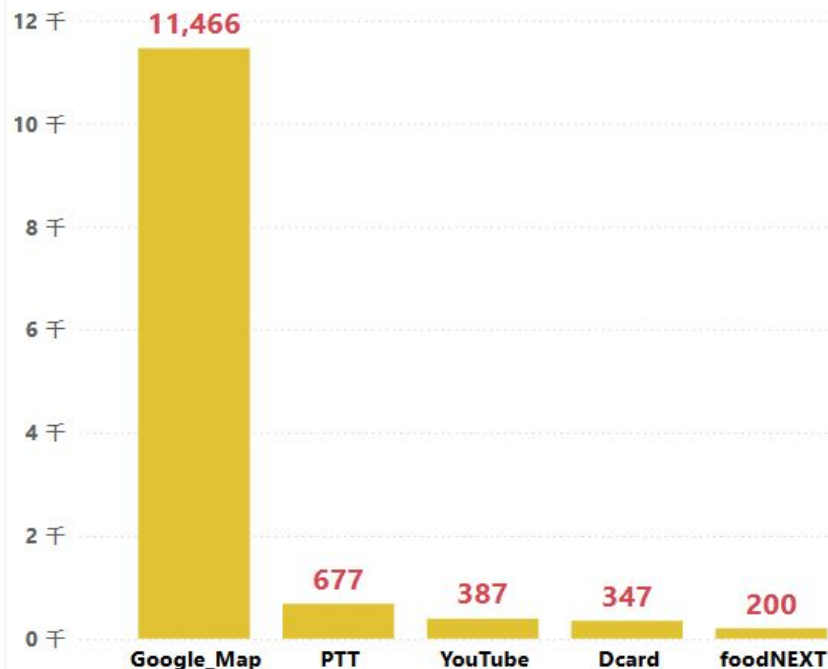
最新資料

2025年2月19日

蒐集各平台資料量佔比



各平台資料量



評分平台數

4

資料總數

12.877 千

最早資料

2021年1月6日

最新資料

2025年2月13日

Cama

Louisa

取消...

2021

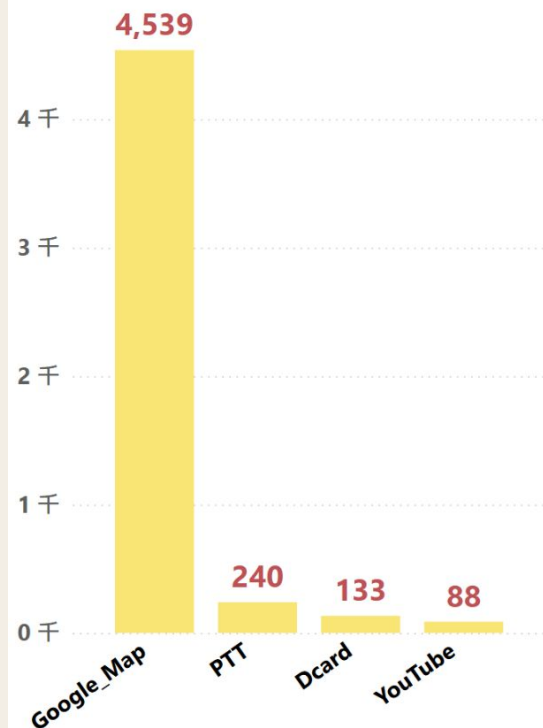
2022

2023

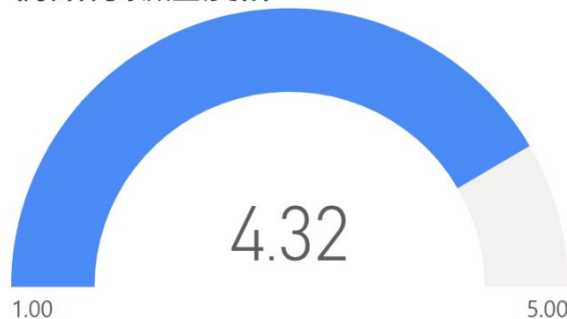
2024

2025

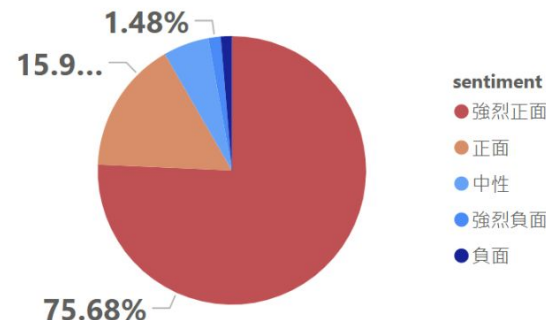
各平台資料量



情緒分數溫度計



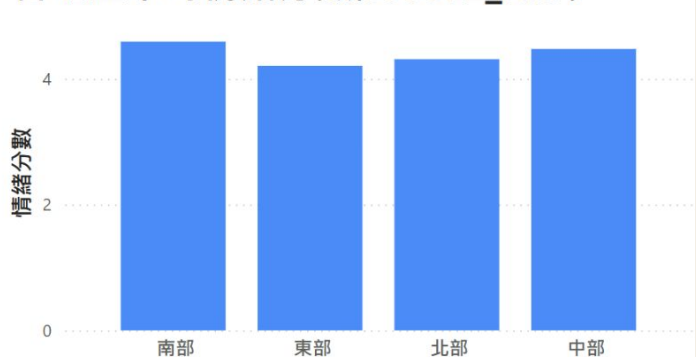
情緒分析等級占比



各年份平均情緒分數



各地區平均情緒分數(GOOGLE_MAP)



評分平台數 資料總數 最早資料 最新資料

4

12.877 千

2021年1月6日

2025年2月13日

Cama

Louisa

取消...

2021

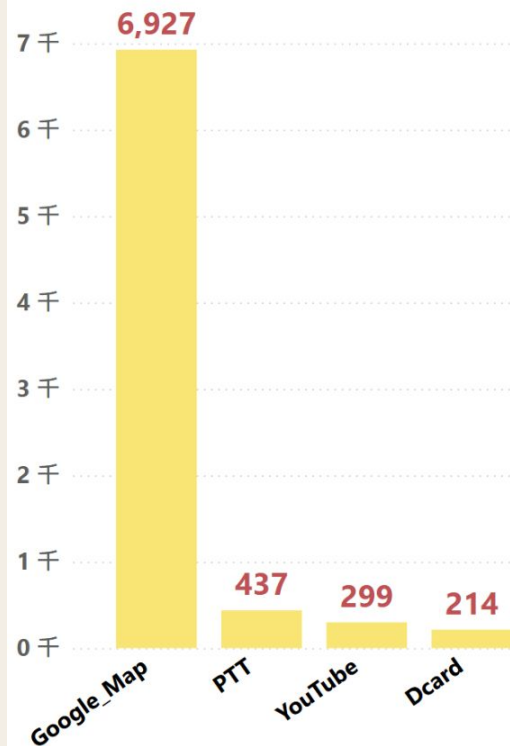
2022

2023

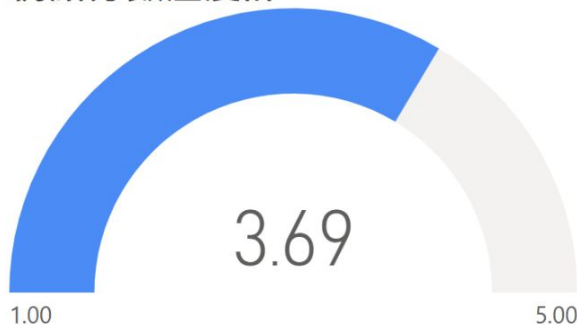
2024

2025

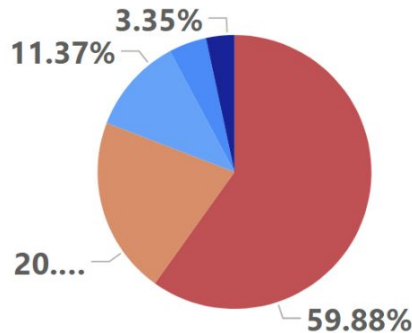
各平台資料量



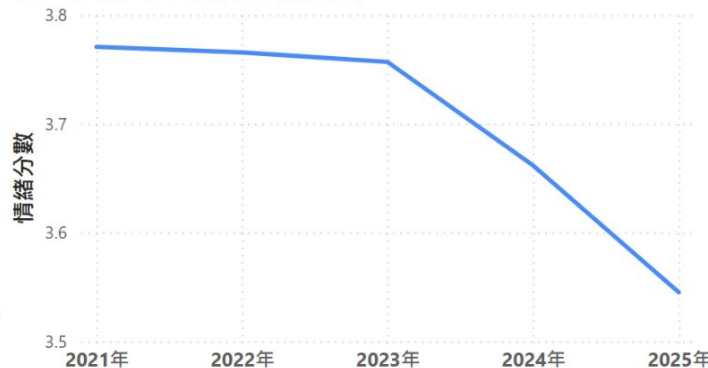
情緒分數溫度計



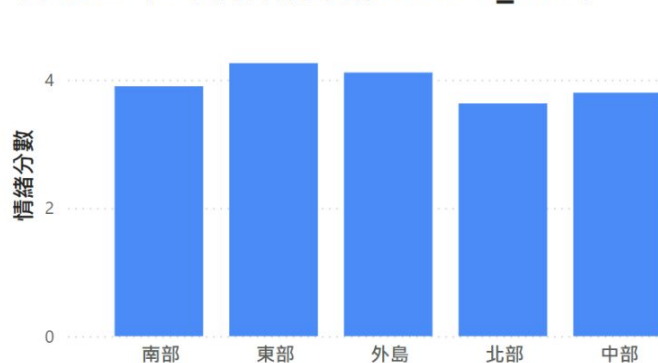
情緒分析等級占比



各年份平均情緒分數



各地區平均情緒分數(GOOGLE_MAP)



過程中碰到的問題及解決方式

前期

- 資料蒐集
- 爬蟲設計

Q: 爬蟲目標不明確

A: 決定以消費者評論為主

過程中碰到的問題及解決方式

中期

- 連結雲端
- 資料處理
- 資料庫設計

Q: 雲端連地端時, 金鑰設置變動?

A: 重開新使用者

Q: 各平台資料不同, 如何統整分析?

A: 核對各 table 欄位, 找出同質性

(例如每個表一定都有 content)

並研究可以區分五等級的語意分析

過程中碰到的問題及解決方式

後期

- 資料整合
- 資料呈現

Q: 視覺化工具選擇？

A: 曾嘗試 Oracle、python 程式，
最終選擇連線最穩定的 PowerBI

未來展望

- 爬取更多網站, 讓分析數據更完整
- 將視覺化分析圖表架設在 API swagger網頁
- 調整語意分析參數、計算誤差率
- 以文字雲圖表呈現關鍵字
- GCP自動備份設置

A background image showing three hands holding coffee cups in a toast. The top hand holds a white cup with latte art. The bottom-left hand holds a dark cup with ice. The bottom-right hand holds a white cup with latte art. The text '謝謝聆聽' is overlaid in the center.

謝謝聆聽