

程式設計與資料科學導論

Final Project - Google map analysis

組名：我們坐在隔壁

B07901069 電機三 劉奇聖
B09902102 資工一 陳冠辰
B06901046 電機四 葉曜德
B06901187 電機四 韓承霖

研究動機

我們常常使用 google map, 除了使用導航的功能之外, 有時也會參考上面的評分和評論幫助我們認識地點。我們隱約注意到地點的種類似乎會影響到評分、評論內容用詞似乎會隨著評分高低有所不同等現象。因此想藉由分析 google map 評論、評分、店家種類三者的關係驗證我們的猜想。

研究目標

- 不同類型的地標在 Google Map 上評分的分佈情形
- 不同類型地標的評論重點詞
- 評分高低的評論重點詞差異
- 觀察 Google Map 上整體評論重點詞

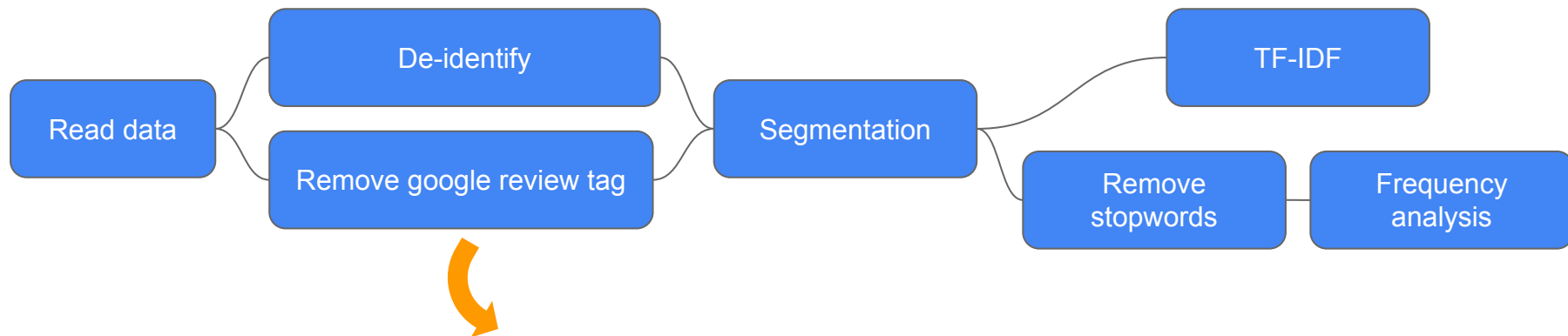
Workflow



資料取得

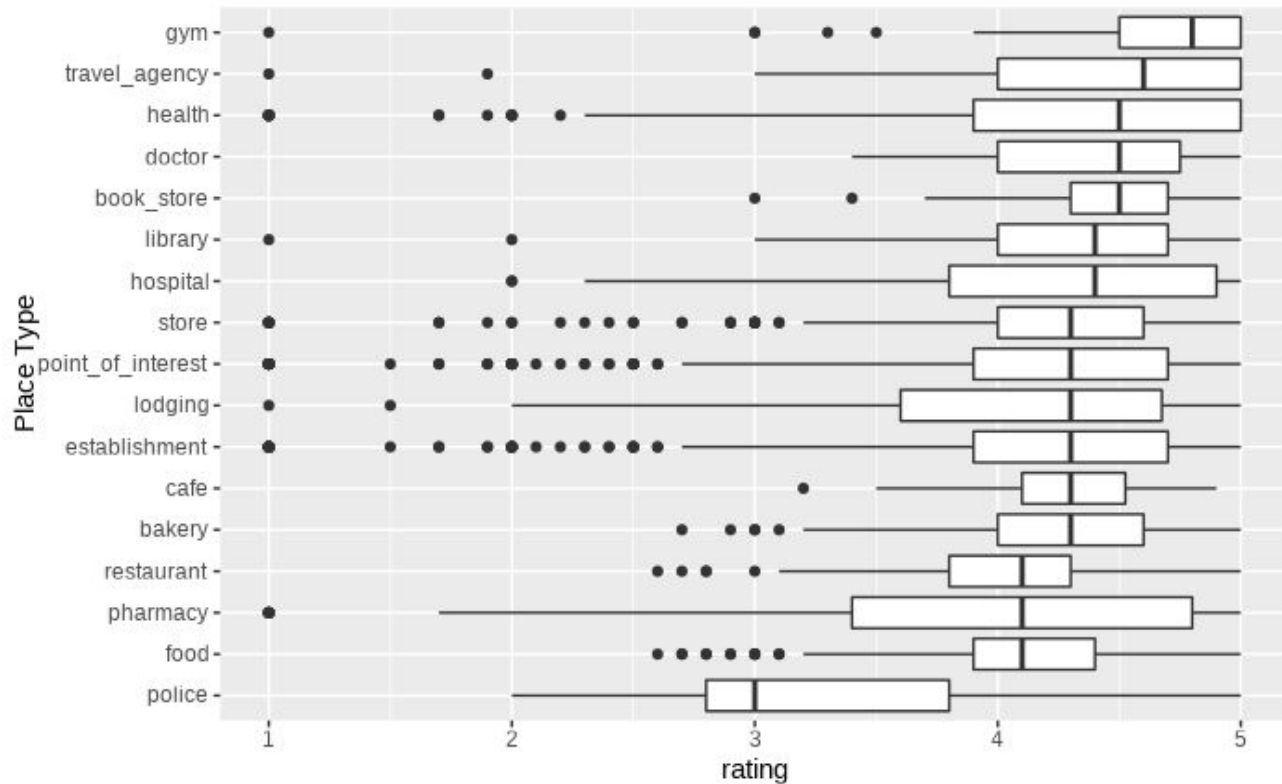
- 首先使用 Google Place API 裡的 Nearby search 抓出台大附近的各種地標資訊約 1000 筆, 裡面包含名稱、評分、地標種類等資訊
- 使用 Google Place API 裡的 Place Details Request 得到各地點在 Google map 上的網址
- 使用 Python 的 Selenium 模擬瀏覽器捲動將各地點的評論爬取下來, 以中文評論為主, 每個地點最多爬取約 400 筆評論, 總評論數量約 86000 筆

資料處理流程



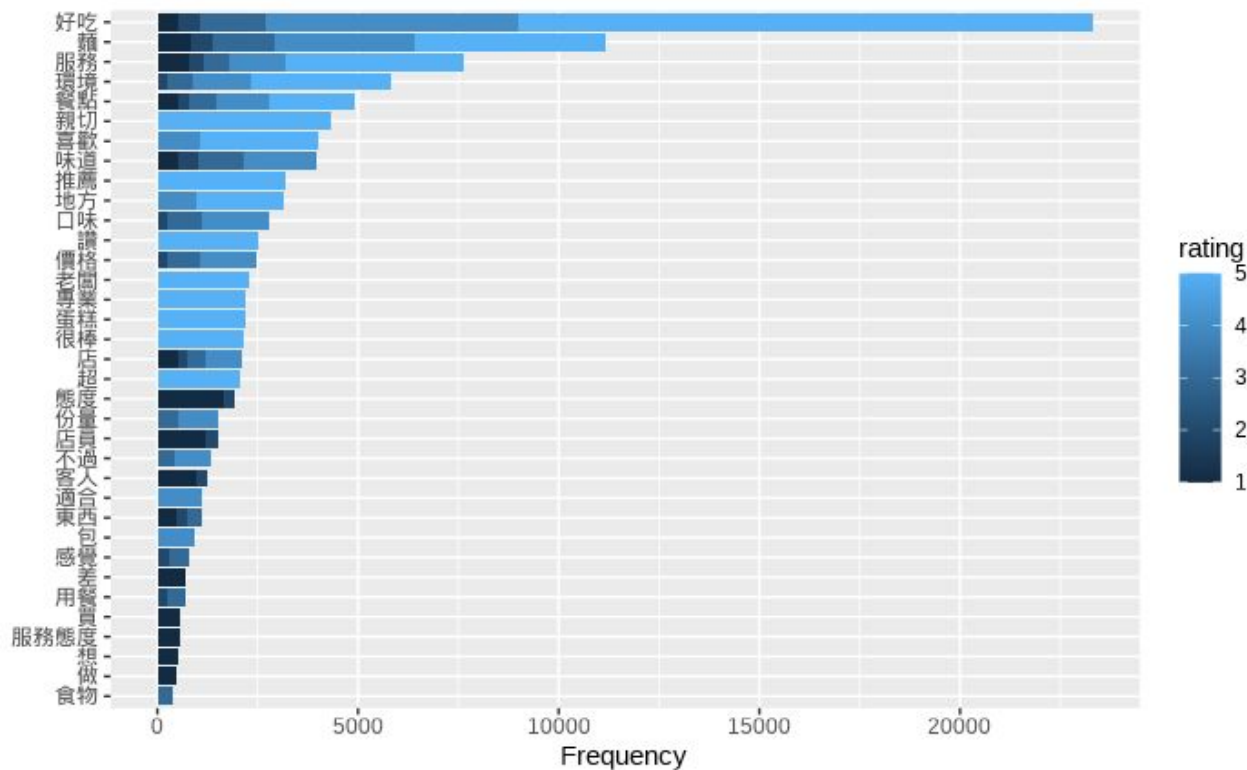
4	(由 Google 提供翻譯) 狹窄的小巷裡有一位藝術家的工作室 風景很美, 值得推薦 (原始評論) 細く入り組んだ路地に芸術家
4	(由 Google 提供翻譯) 古色古香 (原始評論) 古色古香
5	(由 Google 提供翻譯) 寧靜和神聖 (原始評論) Serene and divine
4	(由 Google 提供翻譯) 五點多就關門了, 沒來得及進去。門外有昭和年間的石欄桿。可見歷史悠久。(原始評論) 五點多就
4	(由 Google 提供翻譯) 小而巧 (原始評論) 小而巧
4	(由 Google 提供翻譯) 古色古香 (原始評論) 古色古香
3	(由 Google 提供翻譯) 拍照的好地方 (原始評論) 拍照的好地方
4	(由 Google 提供翻譯) 可愛的藝術家村其中一個歷史悠久的村莊改造而成。由古老的房屋、藝術家工作室和小咖啡館組成。

結果：地標類型的評分分佈

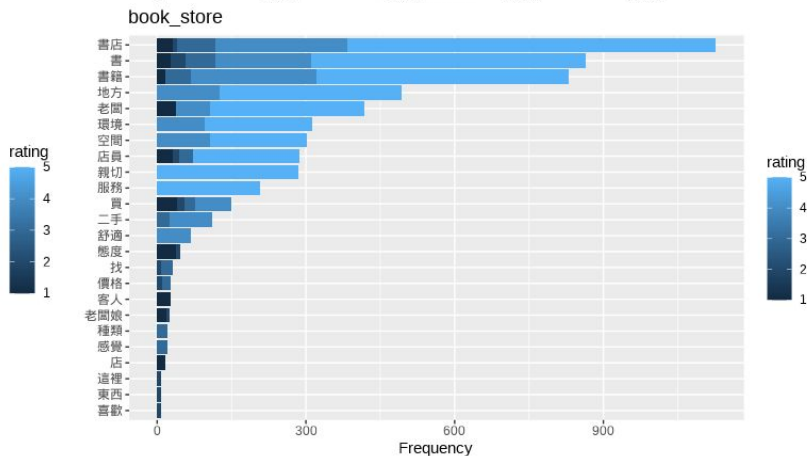
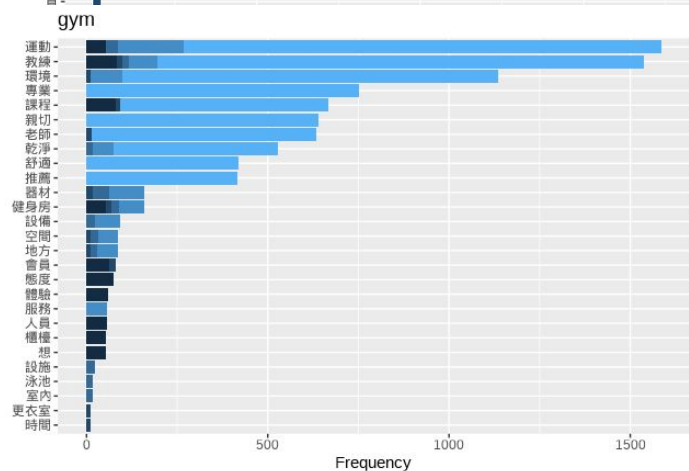
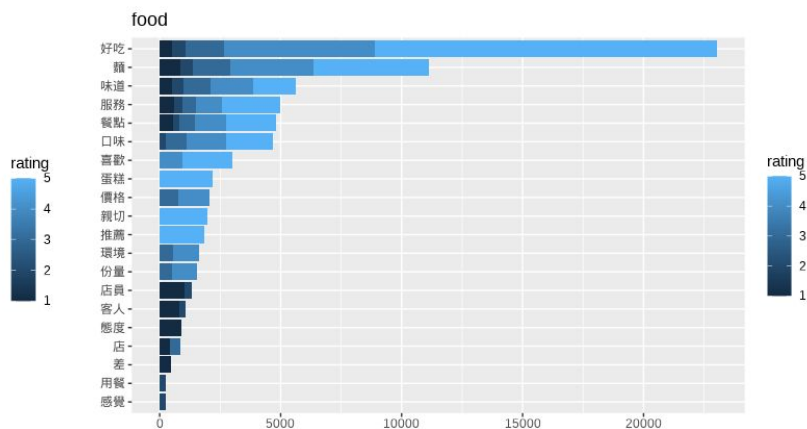
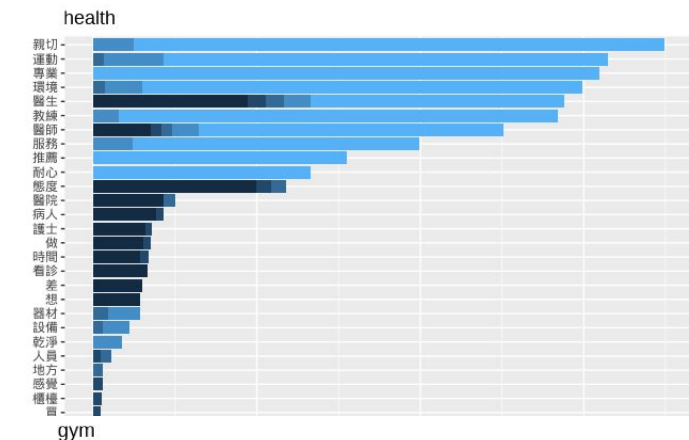


可以看出地點種類確實對評分有所影響，如警察局的評分明顯偏低，休閒場所如體育館、旅遊社等評分偏高。

結果：整體 Google Map 評論的重點詞

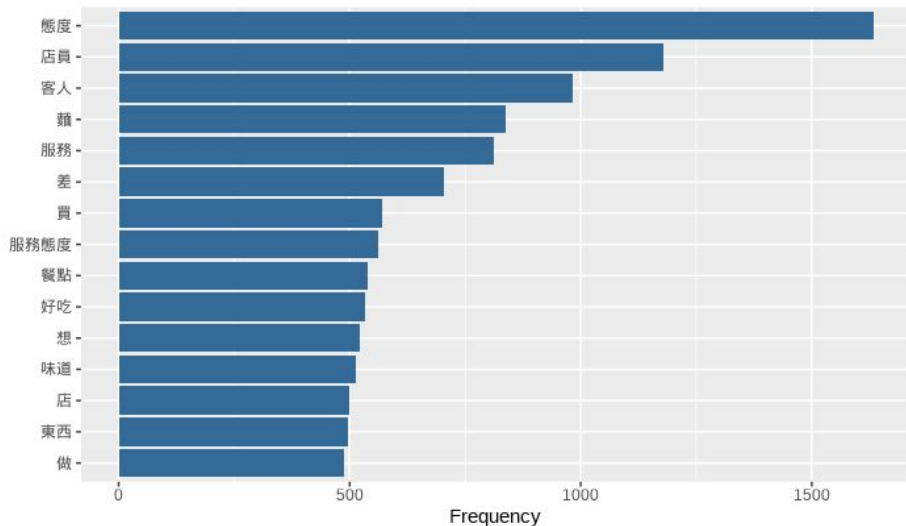


結果:不同地標類別評論重點詞

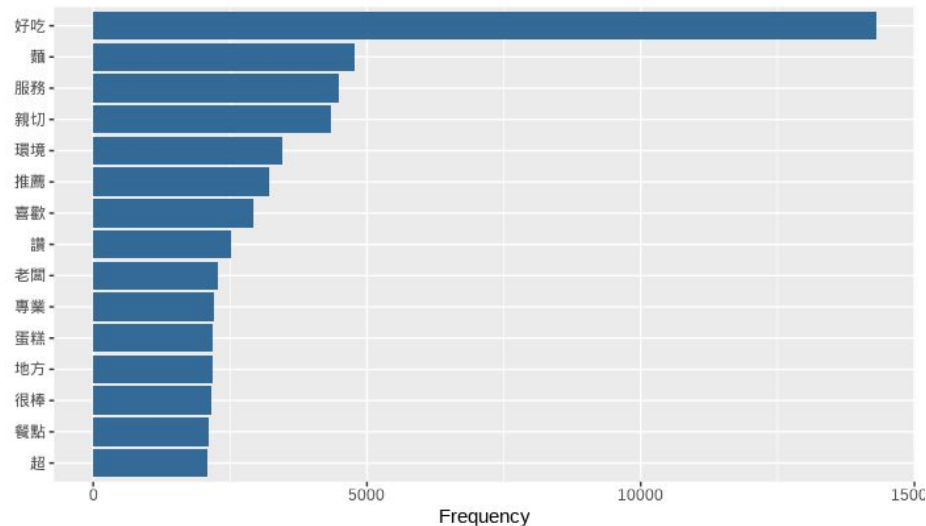


結果：不同評分的評論重點詞

Rating = 1



Rating = 5



程式碼連結

<https://github.com/MortalHappiness/rlads2021Spring-Final>

網頁 Demo



不同地標類別評論重點詞分析 - "health"

This horizontal bar chart displays the importance of various words related to the category 'health'. The y-axis lists 8 words, and the x-axis shows the 'tf_idf' score from 0 to 0.1. The bars are blue. The words are ordered by their importance, with '親切' (Kind) having the highest score at approximately 0.095.

Words in TfIdf ordered	tf_idf
親切	0.095
專業	0.090
運動	0.085
環境	0.080
醫生	0.078
教練	0.075
醫師	0.070
服務	0.065

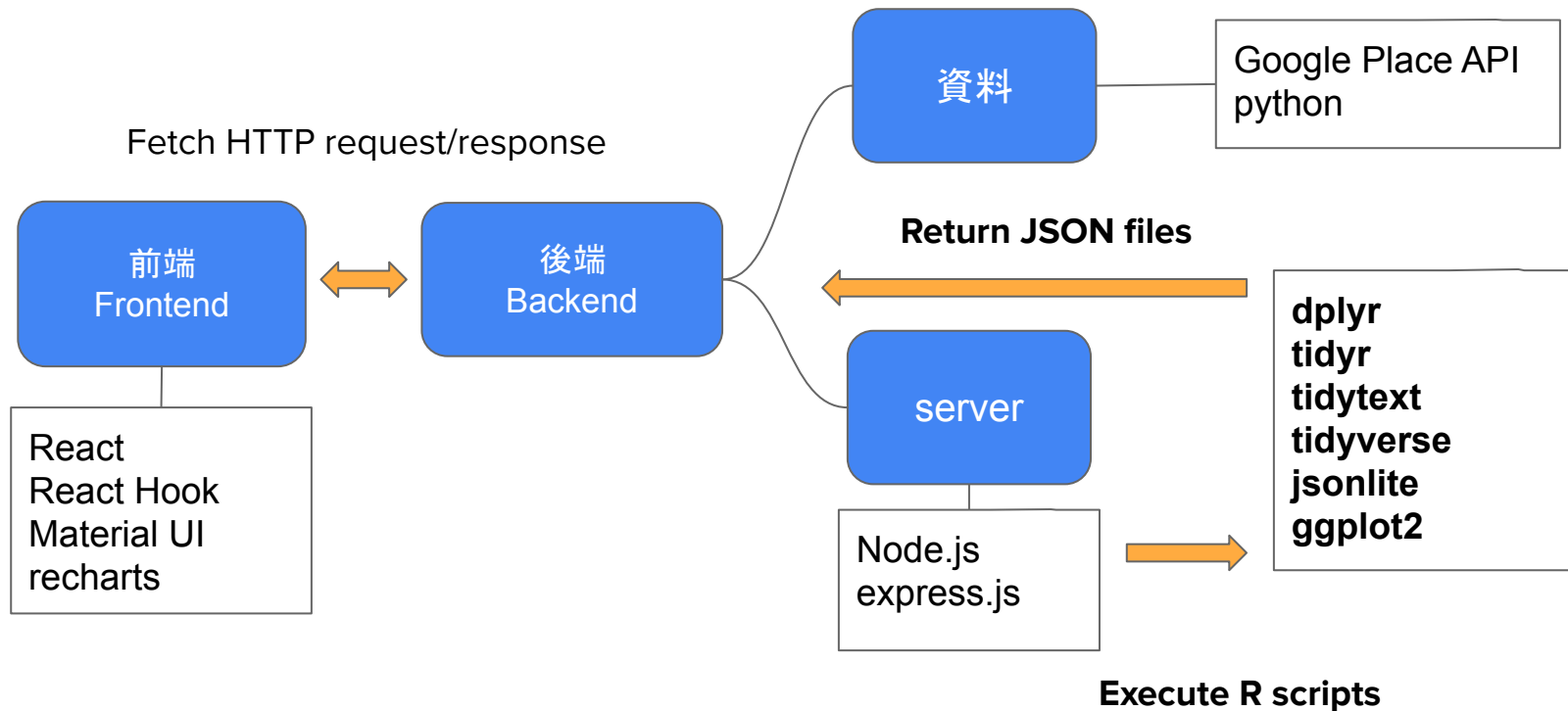
Input Field

Number
15

Type
health

SUBMIT

套件、**API**與架構



分工表

姓名	負責內容	投入程度 (1 ~ 6)
劉奇聖	網頁爬蟲、程式碼整理、資料前處理、不同評分的評論分析、不同種類的評論分析、投影片製作、前端畫面微調	6
陳冠辰	資料前處理、資料分析、作圖、投影片與報告	5
葉曜德	資料後處理、網頁前後端架設、部分資料分析與作圖	5
韓承霖	整理投影片與報告	3