

113356002 資管碩一 陳薇亘 資料模式 Midterm Project

1. Introduction

本次分析目標旨在結合Google公開資料庫資料，了解廣告投放的轉換次數，與廣告的瀏覽量、點擊率以及搜尋趨勢的關係；並利用機器學習的方式，建立監督式預測模型，對於廣告投放轉換率進行預測；最終透過資料儀錶板的方式提供資訊。

下放列出本次分析所使用的資料來源及說明，詳細清理、合併方式請見第二章說明。

Google國際搜尋趨勢 (Google Trends Data)

- 資料來源：Google公開資料庫 (`google_trends.international_top_rising_terms`)
- 表單資訊：提供各國搜尋關鍵字的歷史數據，並以熱門分數(score)表示該關鍵字的熱門程度。

GA網站瀏覽資料 (Google Analytics Data)

- 資料來源：Google公開資料庫 (`google_analytics_sample.ga_sessions_20170801`)
- 表單資訊：提供頁面的瀏覽量(pageviews)與交易量(transactions)的相關資訊。

Google Ads地理目標資料 (Google Ads Geotargets)

- 資料來源：Google公開資料庫 (`google_ads.geotargets`)
- 表單資訊：Google依照地理位置所設定目標受眾的相關資訊，包含區域名稱、區域類型(國家、城市或區域等等)，以及對應到的國家。

Google Ads地理對應資料 (Google Ads Geo Mapping)

- 資料來源：Google公開資料庫 (`ads_geo_criteria_mapping`)
- 表單資訊：詳細列出各個行銷投放準則，對應到各層級的資料；各層級以小至大依序為城市、區域、國家、洲。

客戶廣告行為紀錄 (Customer Interaction Logs)

- 資料來源：本分析取得的相關外部表單(`customer_logs.csv` 上傳後連結至專案資料庫)
- 表單資訊：用戶對於廣告的行為資訊，包含點擊、瀏覽與購買。

廣告成效資料 (Ad Performance Data)

- 資料來源：本分析取得的相關外部表單(`ad_performance_logs.JSON` 上傳後連結至專案資料庫)
- 表單資訊：一共17支廣告的成效資訊，包含廣告的出現、點擊與轉換次數。

2. Data Preparation and Intergration

本次使用的資料表在紀錄的尺度上不盡相同，在單位尺度方面，本分析會先以「國家」為單位，後再以「廣告」為單位進行合併。另外值得注意的是，針對時間尺度，受限於現有的資料特性，會忽略時間尺度做討論。接下來將說明如何清理、串接各個資料表單，並展示合併後的結構。

Step1. 清理Google國際搜尋趨勢

- 原始資料過於龐大，僅隨機抽取0.1%的資料使用
- score欄位缺值很多，考量如果不處理後續模型演算資料點會直接被刪除，因此以平均數填補該值，並新增一個欄位名為「score_null」代表原始資料score欄位是否為缺失值
- 利用MinMax的方式將興趣分數轉到0-1之間，值越大代表該詞越熱門

整併後欄位：country_code, country_name, term, WEEK, score, score_null

Step2. 清理GA網站瀏覽資料

以國家為單位加總瀏覽量及交易量

整併後欄位：country, pageviews, transactions

Step3. 整合Google Ads地理資料

- 合併「地理目標資料」與「地理對應資料」
- 以「地理對應資料」中的 `target_country_region` 作為國家名稱
- 計算每一個國家對應到的目標區域數與目標城市數；因為同一個區域可能出現在很多個準則中，因此使用DISTINCT的方式去重後計算。

整併後欄位：country_code, country_name, target_city_num, target_region_num

Step4. 整合廣告投放資料

- 合併「客戶廣告行為紀錄」與「廣告成效資料」
- 考量客戶行為紀錄中，除了點擊之外，瀏覽(view)以及購買(purchase)也是重要資訊，因此以廣告(ad_id)為單位，計算每一支廣告在各國的點擊量、瀏覽量以及購買量

整併後欄位：ad_id, country, clicks, impressions, click_num, view_num, purchase_num

使用國家名稱或代號串接上述整理後的表單，最終表單包含以下資訊：

欄位名稱	資料型態	說明
ad_id	STRING	廣告編號
ad_performance	RECORD	廣告成效
ad_performance.region_id	STRING	廣告投放區域

ad_performance.impressions	INTERGER	廣告出現次數
ad_performance.clicks	INTERGER	廣告點擊次數
ad_performance.coversions	INTEGER	廣告轉換次數
country_code	STRING	國家代號
country_name	STRING	國家名稱
term	STRING	關鍵字內容
week	DATE	關鍵字所屬週次
score	FLOAT	熱門分數
score_null	INTEGER	熱門分數缺值與否
click_num	INTEGER	各國廣告點擊總數
view_num	INTEGER	各國廣告瀏覽總數
purchase_num	INTEGER	各國廣告交易總數
CTR	FLOAT	各國點擊率
pageviews	INTEGER	各國頁面瀏覽總數
transactions	INTEGER	各國頁面交易總數

3. Key Insights

3.1. 搜尋趨勢熱門分數缺失值

如前述，各國搜尋趨勢的熱門分數在原始資料中有缺失值存在，由下表可以發現資料中熱門分數為缺失值的比例高達八成以上。

熱門分數缺失值	數量	佔比
是	166,440	80.5%
否	40,372	19.5%
整體	206,812	100%

3.2. 國家數量

透過分析可以發現，資料中獨特的國家數量為17個。

3.3. 各國搜尋趨勢

由於資料中搜尋趨勢是按週次產生，為避免結果受到時間因素而失去解讀意義，下表會篩選最新一期(2024-12-01)的資料，並顯示各國搜尋趨勢前5名。值得注意的是，由於資料經過隨機抽樣，因此每個國家在2024-12-01的資料數量不一致。下表呈現「日本」在2024-12-01的搜尋趨勢前5名：

TOP of Japan	term	score (0-1)
1	かつや 感謝 祭	1.0
2	流行 語 2024	1.0
3	モンテディオ 山形	0.848
4	ケンタッキー 福袋	0.798
5	川島 如 恵 留	0.737

3.4. 變數相關性

本小節將分析熱門分數(score)、廣告轉換次數(ad_performance.coverversions)、各國頁面瀏覽總數(pageviews)及各國頁面交易總數(transactions)之間的關係。由下表可以發現，變數關係的相關性並不高。

A	B	Correlation of A, B
熱門分數	廣告轉換次數	-0.0421
熱門分數	各國頁面瀏覽總數	-0.0269
熱門分數	各國頁面交易總數	0.0086
廣告轉換次數	各國頁面瀏覽總數	-0.0607
廣告轉換次數	各國頁面交易總數	0.0000
各國頁面瀏覽總數	個國頁面交易總數	-0.0811

3.5. 各廣告點擊率

透過計算各個廣告的點擊率(CTR)，可以發現資料中點擊率最高的國家依序為AD018、AD001、AD004。以廣告AD018為例，每個廣告出現一次平均可獲得7.5次的點擊數，約為整體的6倍。

廣告別	點擊率
AD018	7.5
AD001	3.0
AD004	2.7
整體平均	1.2

3.6. 各國廣告點擊率

透過計算各國平均點擊率(Mean of CTR)，可以發現資料中點擊率最高的國家依序為巴西、挪威及瑞士。以巴西為例，每個廣告出現一次平均可獲得7.5次的點擊數，約為整體平均的6倍。值得注意的是，由於目前資料中，巴西、挪威及瑞士與廣告的對應關係屬於1對1對應，因此分析的結果會等價於各廣告點擊率。

國家	平均點擊率
巴西(Brazil)	7.5
挪威(Norway)	3.0
瑞士(Switzerland)	2.7
整體平均	1.2

4. Predictive Models

本研究利用監督式機器學習的線性迴歸模型，利用現有變數對目標變數 - 廣告投放轉換率進行預測，並將前述經過整理的資料，依照8:2分為訓練集與測試集，並利用MAE、RMSE及 R^2 對於模型效能進行評估。

資料	筆數	佔比
訓練資料集	165,450	80%
測試資料集	41362	20%
整體資料	206,812	100%

本研究共嘗試兩個模型：

Simple Model

Simple Model使用的變數量及資訊複雜度較低，其對於轉換率的變異解釋能力為14.4%。

Complex Model

基於Simple Model的基礎，本研究額外加入三個新變數訓練Comple Model，分別為score_null、view_num以及purchase_num，其對於轉換率的變異解釋能力為32.5%，為Simple Model的兩倍。

目標變數(Y) = conversions (轉換率)

模型	變數(x)	MAE (測試集)	RMASE (測試集)	R^2 (測試集)

Simple Model	<ul style="list-style-type: none"> - score - CTR - click_num 	253.992	280.111	0.144
Complex Model	<ul style="list-style-type: none"> - score - score_null(新增) - CTR - click_num - view_num(新增) - - purchase_num(新增) 	210.446	248.744	0.325

5. DASHBOARD SUMMARY

本次分析利用Looker Studio提供儀錶板功能供使用者使用。

頁面1. Region Insights

頁面1提供使用者區域性的分析，右上角顯示各區域對應的平均點擊率，下方圖片則顯示各國的搜尋趨勢，柱狀圖的高度越高代表score數值越大。使用者可以透過左邊中間的篩選項選擇想要查看的ad_id、term以及country；除此之外，使用者也可以點擊各個顏色的條狀圖，右鍵選擇下鑽功能，就可以進一步分析各條對應到的時間資訊。

使用者可透過篩選目標國家，了解各國的搜尋趨勢，讓廣告能與現有的趨勢做結合。

圖1. Region_Insights頁面

Region Insights

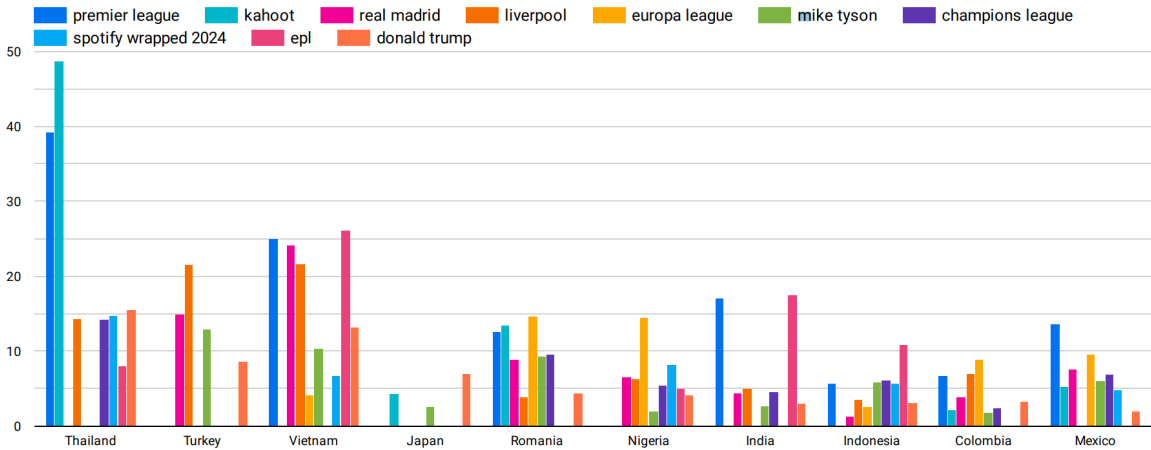
ad_id ▾

term ▾

country_name ▾

Region vs CTR	
Region_id	AVG_CTR ▾
R01	4.31
R06	2.25
R03	1.46
R05	0.47
R04	0.46
R10	0.36
R09	0.30
R07	0.16

1 - 8 / 8 < >



頁面2. Ad_Performance

頁面2提供使用者查詢廣告的成效表現，右上角顯示的是各個時間點對應到的平均點擊率。下方呈現的則是本分析建置的模型(Simple Model及Complex Model)在預測各廣告轉換率的絕對誤差。

圖2. Ad_Performance頁面

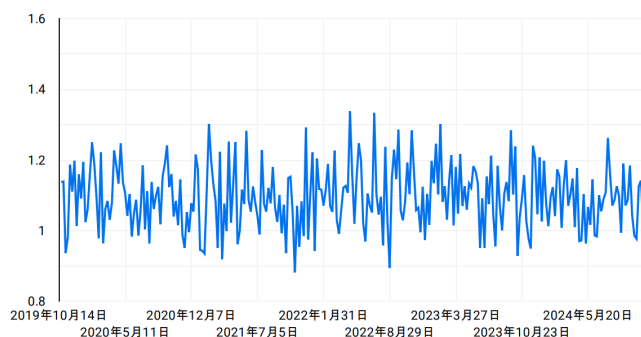
Ad Performance

ad_id ▾

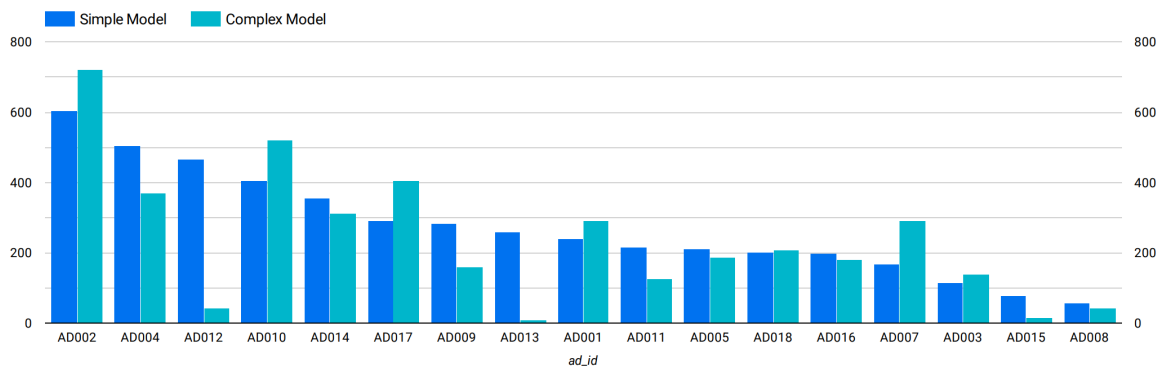
term ▾

country_name ▾

點擊率變化趨勢圖



絕對誤差_轉換率



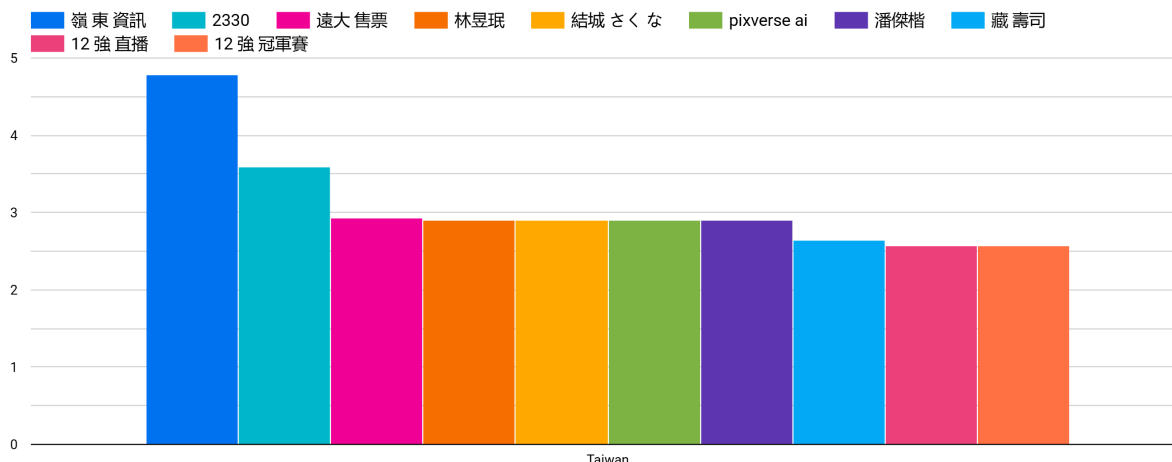
本分析利用頁面1針對巴西、台灣、以及日本三個國家，分析時下的搜尋趨勢，並提供廣告商投放建議：

巴西

篩選巴西的搜尋趨勢前10名，可以發現很多都與足球賽事有關，如「corinthians x」、「flamengo x」、「atlético-go x atlético-mg」等皆是巴西足球聯賽的參賽隊伍，因此廣告商可以考慮跟足球賽事合作，吸引民眾點擊相關廣告。

台灣

篩選台灣的搜尋趨勢排名前10，有將近一半與世界盃棒球賽有關，包含林昱珉、潘傑凱、12強直播、12強冠軍賽等，廣告商可以考慮與這些棒球選手合作代言，或在廣告的內容中納入「慶祝台灣棒球奪冠」等內容，增加民眾對於廣告的好奇心，進而增加廣告的效益。



日本

篩選日本的搜尋區是排名，發現第一名為JRA(日本中央競馬會)，顯示日本民眾近期較為關注競馬，可考慮與賽事合作，進行聯名代言的廣告投放等。

6. Conclusion and Recommendations

6.1. 實務建議

本次研究發現，大多數國家的搜尋趨勢都與運動賽事有關，如巴西的足球、台灣的棒球以及日本的競馬，顯示運動已成為跨國籍、跨文化的大眾休閒娛樂；因此廣告商在進行廣告投放時，可以考慮與運動賽事、選手進行結合，以提高民眾的關注度。另外值得注意的是，現有的廣告投放觸及的僅有17個國家，且多集中在歐美地區，未來可考慮增加廣告預算，並對目前還未觸及的國家、區域進行廣告投放。

6.2. 未來分析建議

本次研究考量現有資料的欄位限制、資料表間尺度的差異，導致整合後的資料較難給出具體的建議，且在模型預測的表現能力也還有改善空間。建議未來可針對下列幾點進行改善：

考量搜尋趨勢與廣告的契合程度

由於現有的資料並未有廣告本身的相關特徵，如廣告本身的訴求、相關的產品、廣告本身的文字資訊等等，因此較難與搜尋趨勢進行結合。未來建議可以蒐集相關數據，分析廣告本身內容與搜尋趨勢的契合程度，是否能有效增加廣告的轉換率，以提供決策參考。

更細緻的資料分析尺度

受限於資料本身的欄位，目前僅能夠以「國家」為單位合併各表，但分析過程中也發現，Google Ad的欄位有詳細到城市、區域、省等等，建議未來可以想辦法結合其他外部表單，縮小資料分析的顆粒度，以洞察更細緻的資訊。

考量時間因素

本次分析並未考量時間因素，然而搜尋趨勢應該是瞬息萬變的，建議未來應納入時間因素進行分析，以擷取對於當下決策最有幫助的資訊。

附件

專案關閉截圖

關閉「final-113356002」專案

關閉專案之後：

- 專案關閉後的一段時間內可能會持續產生費用。為避免產生非預期的費用，請先停用計費功能再關閉專案。
- 會在幾小時內停止提供流量。
- 您會失去這項專案的存取權。
- 專案擁有者會收到通知，並能在 30 天內取消刪除專案。
- 專案會排定於 30 天後刪除，不過某些資源可能較早。

[進一步瞭解關閉程序。](#) [🔗](#)

如要關閉專案「final-113356002」，請輸入專案 ID：**final-113356002**

專案 ID *

final-113356002

取消

仍要關閉

帳單截圖

總費用 (2024/12/5至
2024/12/11)



最近 7 天

當前月分

費用
\$0.02 — 使用的抵免額
\$0.02 =

[查看詳細資料](#)

總費用
\$0.00

預測總費用

累積足夠的使用時間之後，
即可查看這項資料

\$1

\$0.50