

Midterm Project

陳泓霖

113356033

Project close & Billing

關閉「looker-assignment-113356033」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「looker-assignment-113356033」，請輸入專案 ID：**looker-assignment-113356033**

專案 ID *
looker-assignment-113356033

[取消](#) [仍要關閉](#)

總覽 [免費試用帳戶](#) [管理帳戶](#)

[帳單帳戶總覽](#) [付款總覽](#)

總費用 (2024/12/1 至 2024/12/11) ⓘ

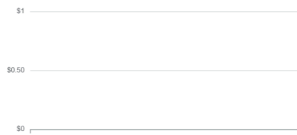
最近 7 天 [當前月分](#)



→ 前往「報表」頁面查看詳細資料

用量最多的專案

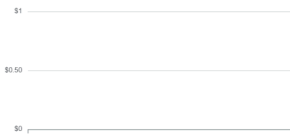
2023/12/1 至 2024/12/31



→ 查看報表

常用服務

2023/12/1 至 2024/12/31



→ 查看報表

建立預算快訊 ⓘ

設定帳單帳戶的每月預算。費用接近及超過預算金額時，您會收到電子郵件通知。

\$

[建立](#)

→ 查看預算與警告

帳單帳戶健康狀態檢查

您可以查看帳戶的健康狀態檢查結果，藉此預防常見的帳單相關問題，以及採用我們提供的最佳做法建議。[瞭解詳情](#)

0 1 1

→ 查看所有健康狀態檢查

免費試用抵免額



免費試用期間不需要支付任何費用。免費試用期結束之後，除非升級為 Cloud Billing 付費帳戶，否則您名下的所有資源都會停止，而系統也不會向您收費。

請注意：免費試用期無法暫停或延長。請您用所有可用的抵免額，或是免費試用期已到期後，免費試用期將會結束。(以先發生者為準)。

[啟用](#) [瞭解詳情](#)

1. Introduction

目標：

作為一名資料工程師，受到一家全球廣告公司的委託，製作一份對於如何優化廣告策略的報告。

報告中的內容應該包括：

- 1. 識別具有高度廣告潛力的地區與主題
- 2. 分析廣告表現以識別各地區與廣告之表現
- 3. 建構預測模型
- 4. 使用專業的形式呈現可行的全球策略

資料集介紹：

1.Google Trends Data

記錄了各國及區域在特定週期內的熱門關鍵字、排名、得分與百分比增長，幫助分析全球搜尋趨勢和關鍵字熱度的變化。

International Top Rising Terms

欄位名稱	類型	模式
country_name	STRING	NULLABLE
region_name	STRING	NULLABLE
region_code	STRING	NULLABLE
week	DATE	NULLABLE
rank	INTEGER	NULLABLE
country_code	STRING	NULLABLE
term	STRING	NULLABLE
score	INTEGER	NULLABLE
percent_gain	INTEGER	NULLABLE
refresh_date	DATE	NULLABLE

2.Google Analytics Sample Dataset

包含有關網站訪問者的詳細資訊，包括每次訪問的訪問者 ID、訪問次數、頁面瀏覽量、停留時間、跳出率、交易數量、交易收入等指標，並且包含了來自不同設備、地理區域、來源的資料，用於分析網站流量、轉換以及使用者行為。

ga_sessions_20170801

欄位名稱	類型	模式
visitorid	INTEGER	NULLABLE
visitNumber	INTEGER	NULLABLE
visitid	INTEGER	NULLABLE
visitStartTime	INTEGER	NULLABLE
date	STRING	NULLABLE
totals	RECORD	NULLABLE
visits	INTEGER	NULLABLE
hits	INTEGER	NULLABLE
pageviews	INTEGER	NULLABLE
timeOnSite	INTEGER	NULLABLE
bounces	INTEGER	NULLABLE
transactions	INTEGER	NULLABLE
transactionRevenue	INTEGER	NULLABLE
newVisits	INTEGER	NULLABLE
screenviews	INTEGER	NULLABLE
uniqueScreenviews	INTEGER	NULLABLE
timeOnScreen	INTEGER	NULLABLE
totalTransactionRevenue	INTEGER	NULLABLE
sessionQualityDim	INTEGER	NULLABLE
trafficSource	RECORD	NULLABLE
device	RECORD	NULLABLE
geoNetwork	RECORD	NULLABLE
customDimensions	RECORD	REPEATED
hits	RECORD	REPEATED
fullVisitorid	STRING	NULLABLE
userid	STRING	NULLABLE
clientid	STRING	NULLABLE
channelGrouping	STRING	NULLABLE

3.Google Ads Geotargets Data

地理定位的相關資訊，包括唯一標識的地理區域 ID、英語名稱、完整名稱、上級區域 ID、國家代碼、地理類型（如城市、國家、省份等）以及地區的狀態（如有效或計劃移除），用於地理定位分析和目標設定。

欄位名稱	類型	模式
criteria_id	STRING	NULLABLE
en_name	STRING	NULLABLE
canonical_name	STRING	NULLABLE
parent_id	STRING	NULLABLE
country_code	STRING	NULLABLE
display_feature_type	STRING	NULLABLE
status	STRING	NULLABLE

4.Google Ads Geo Mapping Data

這個資料集將廣告目標的地理位置與相應的地區、城市、國家、次大陸和大陸進行對應，幫助分析和優化廣告的地理定位策略。

欄位名稱	類型	模式
ads_criteria_id	INTEGER	NULLABLE
target_city	STRING	NULLABLE
target_region	STRING	NULLABLE
target_country_region	STRING	NULLABLE
target_subcontinent	STRING	NULLABLE
target_continent	STRING	NULLABLE

5.Customer Interaction Logs

這個資料集包含了不同國家的廣告互動數據，記錄了每次廣告點擊、查看或購買的用戶行為，並標註了相應的時間戳和廣告ID。

欄位名稱	類型	模式
log_id	STRING	NULLABLE
timestamp	TIMESTAMP	NULLABLE
ad_id	STRING	NULLABLE
user_action	STRING	NULLABLE
Country	STRING	NULLABLE

6.Ad Performance Data

不同廣告的表現數據，記錄了每個廣告在不同地區的曝光次數、點擊次數、轉換次數及其相關國家。

欄位名稱	類型	模式
ad_id	STRING	NULLABLE
region_id	STRING	NULLABLE
impressions	INTEGER	NULLABLE
clicks	INTEGER	NULLABLE
conversions	INTEGER	NULLABLE
country	STRING	NULLABLE

2.Data Preparation and Integration

資料處理過程：

1. 將本地端的csv檔案上傳至bucket, 再由bucket至bigquery建立表格

2. 處理Google trends data

由Google trends data透過MOD的方式選取(1/10000)的資料, 因為後續要進行機器學習的資料分析, 透過這樣的方式可以切分訓練資料與測試資料。

並且因為各國家間的資料數量並不平衡, 若是完全隨機選取可能無法選取到資料量較少的國家, 導致進行分析時對資料的了解可能有所偏誤。

為達此目的, 首先使用首先使用ROW_NUMBER()對每個國家的資料進行編碼, 再對row number取餘數, 以同是達到三個目的。

- 1.每個國家的資料皆會出現在表格內
- 2.固定隨機選取的範圍, 以利後續進行分析
- 3.減少資料量, 避免數據過多而無法處理

剔除資料表中score為null的資料, 因為後續進行分析時會使用到score。

使用sklearn套件中的MinMaxScaler對score欄位進行正規化。

3. 整合各表格

利用pandas的merge function對各資料集進行inner join。

- Ads Geotargets和Geo Mapping使用ads_criteria_id和criteria_id進行JOIN
- Trends data和Geo data JOIN完的表格使用country_code進行JOIN
- 最後使用country對所有的資料集進行JOIN

所有資料集整合在同一資料集後, 使用pandas的drop和drop_duplicates function 剔除資料集中重複的欄位, 和此次分析中不需使用到的column

最後將此資料集上傳至bigquery的dataset

以下是整合表格integrated_ad_analysis的schema, 共有58,501,517筆資料

欄位名稱	類型	模式
country_code	STRING	NULLABLE
country_name	STRING	NULLABLE
term	STRING	NULLABLE
score	FLOAT	NULLABLE
ads_criteria_id	INTEGER	NULLABLE
ad_id	STRING	NULLABLE
impressions	INTEGER	NULLABLE
clicks	INTEGER	NULLABLE
conversions	INTEGER	NULLABLE
timestamp	TIMESTAMP	NULLABLE
user_action	STRING	NULLABLE
country	STRING	NULLABLE
pageviews	INTEGER	NULLABLE
transactions	INTEGER	NULLABLE

3.Key Insights

以下是本次我們所進行的一些SQL analysis

1.Regional Trends

Malaysia	jill stein	1.0	1
Malaysia	magnum result	0.929292929292...	2
Malaysia	toto	0.828282828282...	3
Malaysia	barcelona	0.787878787878...	4
Malaysia	epl table	0.787878787878...	4
Mexico	emiliano zapata	0.898989898989...	1
Mexico	buho legal	0.858585858585...	2
Mexico	dof	0.838383838383...	3
Mexico	champions league	0.818181818181...	4
Mexico	seleccion mexicana	0.747474747474...	5

此為部分的bigquery結果，此查詢顯現了各國家排名前5的關鍵字

以此可以了解到在不同的國家目前的流行趨勢

Insight：可透過此排名思考產品能否透過與當地的熱門關鍵字產生連結，製作客製化的廣告，使公司要進行行銷的產品與當地民眾產生共鳴，刺激購買慾望。
且若是廣告中包含當地熱門的關鍵字，也可使廣告的曝光機率提高。

2.Correlations

計算各國家的score, conversions, page views, and transactions彼此間的相關係數
希望能夠藉此了解哪些因素可以提高轉換率和交易數。

	country_name	corr_score_conversions	corr_score_pageviews	corr_score_transactions	corr_conversions_pageviews	corr_conversions_transactions	corr_pageviews_transactions
0	Canada	NaN	-6.400761e-07	NaN	NaN	NaN	NaN
1	United Kingdom	NaN	1.667973e-06	NaN	NaN	NaN	NaN
2	Switzerland	NaN	-2.331277e-07	NaN	NaN	NaN	NaN
3	Japan	NaN	1.058318e-06	NaN	NaN	NaN	NaN
4	Brazil	NaN	1.670072e-06	NaN	NaN	NaN	NaN
5	Malaysia	NaN	3.782041e-07	NaN	NaN	NaN	NaN
6	Mexico	NaN	6.176652e-07	NaN	NaN	NaN	NaN
7	Ukraine	NaN	NaN	NaN	NaN	NaN	NaN
8	Italy	NaN	8.952265e-07	NaN	NaN	NaN	NaN
9	Spain	NaN	-3.987269e-06	NaN	NaN	NaN	NaN
10	Germany	NaN	2.416317e-07	NaN	NaN	NaN	NaN
11	Turkey	NaN	-1.105300e-06	NaN	NaN	NaN	NaN
12	Norway	NaN	8.940248e-16	NaN	NaN	NaN	NaN
13	Thailand	NaN	-5.616654e-07	NaN	NaN	NaN	NaN
14	Indonesia	NaN	-4.272633e-07	NaN	NaN	NaN	NaN
15	Poland	NaN	-1.698643e-07	NaN	NaN	NaN	NaN
16	Netherlands	NaN	4.285094e-07	NaN	NaN	NaN	NaN

可惜的是除了score與page views間的相關係數皆無法計算，且兩者間的相關性也不大

此分析未能找到原先預期所能發現的結果。

3.Ad Performance

使用平均點集率(CTR)進行分析，企業可以識別哪些廣告在各個國家最具吸引力。較高的 CTR 表示該廣告吸引了大量的點擊，可能表明其廣告設計、內容或定向策略有效。

	ad_id	country_name	avg_ctr
0	AD001	Norway	3.030445
1	AD002	United Kingdom	0.567224
2	AD003	Spain	0.360376
3	AD004	Switzerland	2.740538
4	AD005	Malaysia	1.444444
5	AD007	Indonesia	0.830469
6	AD008	Ukraine	1.464676
7	AD009	Turkey	0.572660
8	AD010	Canada	0.166361
9	AD011	Mexico	0.207663
10	AD012	Poland	0.158819
11	AD013	Netherlands	0.051874
12	AD014	Japan	0.434123
13	AD015	Italy	0.272128
14	AD016	Thailand	0.198182
15	AD017	Germany	0.270069
16	AD018	Brazil	7.482578

Insight : CTR的計算為clicks/impressions，部分國家的clicks大於impressions，不確定數據是否有錯誤。若假設此為正常現象，我們可以針對CTR較低的廣告和國家重新擬定該國家的廣告投放策略，因為這些廣告相較來說成效較差。

4.Statistical Approximations

統計目前投放廣告的國家數量

estimated_unique_countries

17

Insight : 目前有投放廣告的國家僅17個，而公開資料集中的trends data出現了42個國家。可以考慮替換目前CTR較低的國家，並根據trend data進行地區客製化的廣告，以期能開發更多陌生市場，增加產品銷售。

4. Predictive Models

模型目的：

本次的模型旨在使用線性迴歸模型預測廣告的轉換率(conversion)

輸入的資料包含

ad_id

country_name

score

ctr

user_action

標籤為

conversion

模型表現：

MAE	RMSE	R ²
252.729	293.541	0.102

MAE(Mean Absolute Error)：測量預測值與實際值的絕對誤差平均值。

RMSE(Root Squared Error)：評估模型預測值與實際值之間的平方差，對於大偏差有更高敏感性。

R²：評估模型解釋資料變異的能力，值越接近 1，模型表現越佳。

此模型的MAE和RMSE數值都偏大，且R²數值也小，變數間的可解釋性低，表現不佳推測為選定的欄位與conversion間的相關性不足，試著加入其他欄位以改進模型表現

模型改進：

訓練資料加入timestamp

MAE	RMSE	R ²
127.468	171.502	0.695

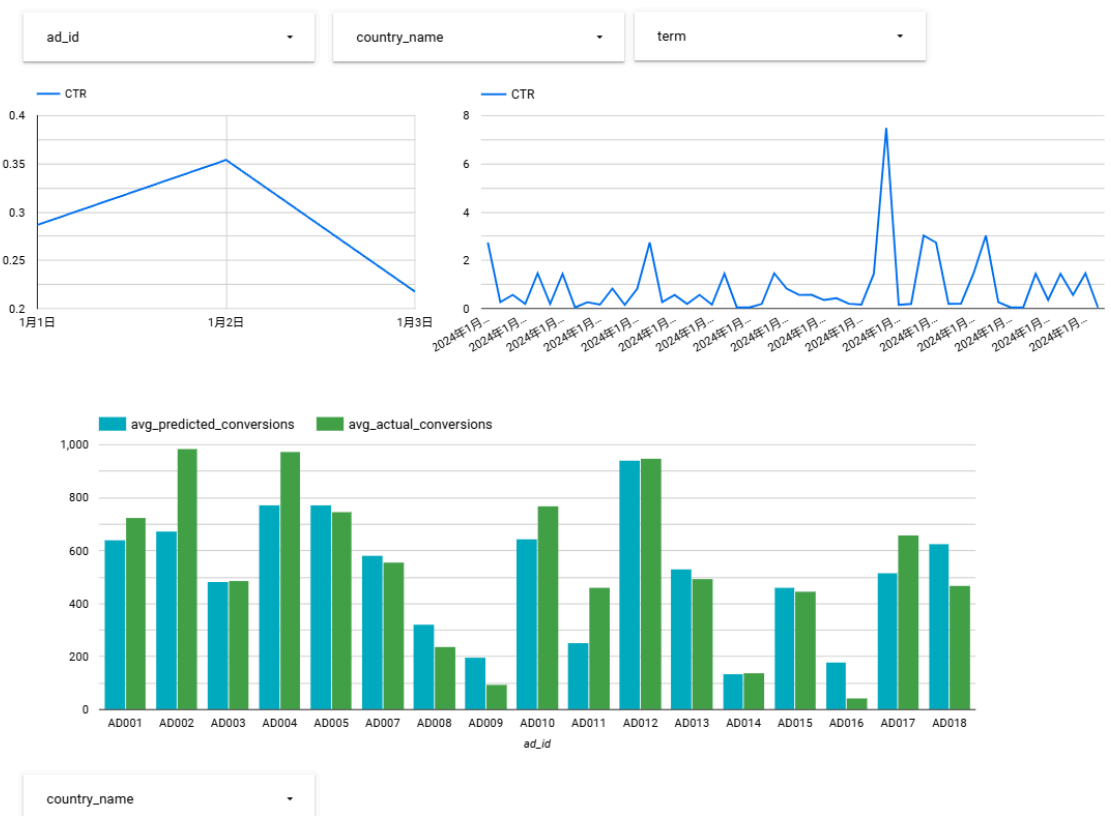
模型的精準度大幅上升：MAE、RMSE數值顯著下降

模型解釋力增強：模型中已有接近7成的變異性可被解釋

可能原因：時間特徵幫助捕捉了廣告轉換鐘的時間模式

例如：產品本身的季節性、廣告的時間效應(促銷或是假期)

5. Dashboard Summary



目前的圖表如何協助決策

1.Bar chart: Top term count per country

每個國家的熱門 Term 分析，幫助優化內容和關鍵字策略。

2.Heatmap: CTR by region

識別 CTR 高的地區，幫助調整廣告投放策略。

3.Line chart: CTR trends over time

了解時間趨勢，分析CTR高的時段之原因，並針對不同時段調整策略。

4.Bar Chart with Error Bars

比較預測值與實際值。

下鑽查詢

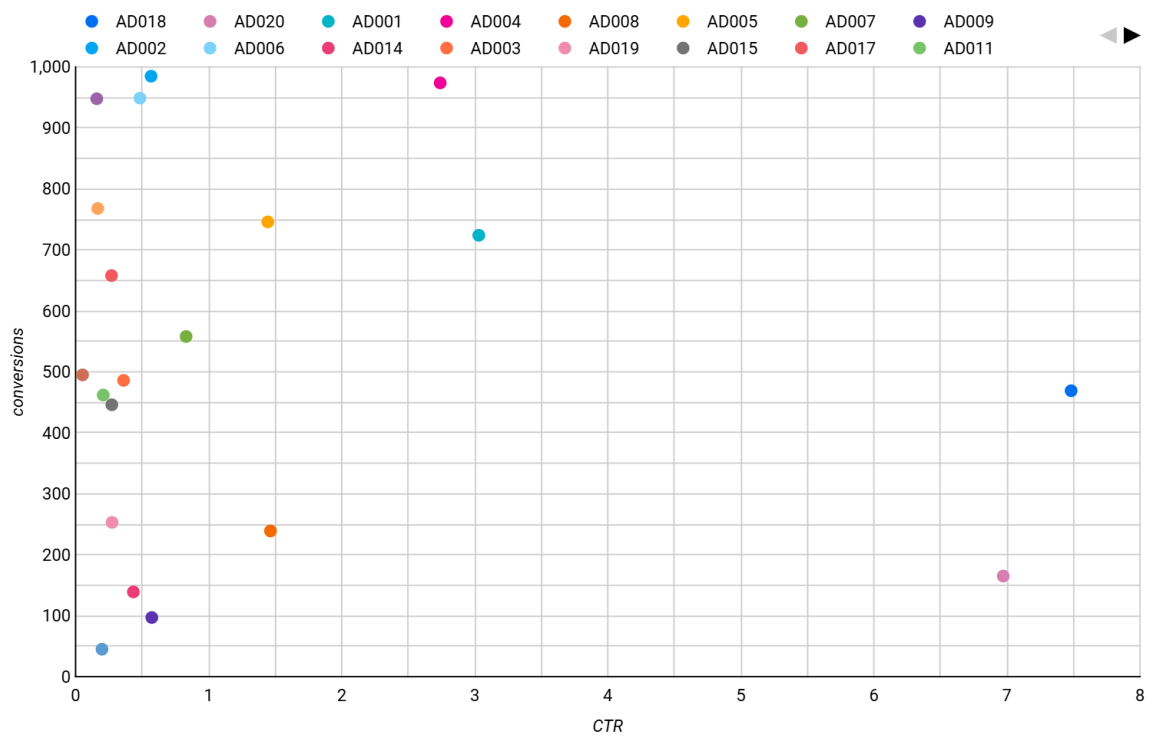


預設啟用下鑽查詢層級

country_name

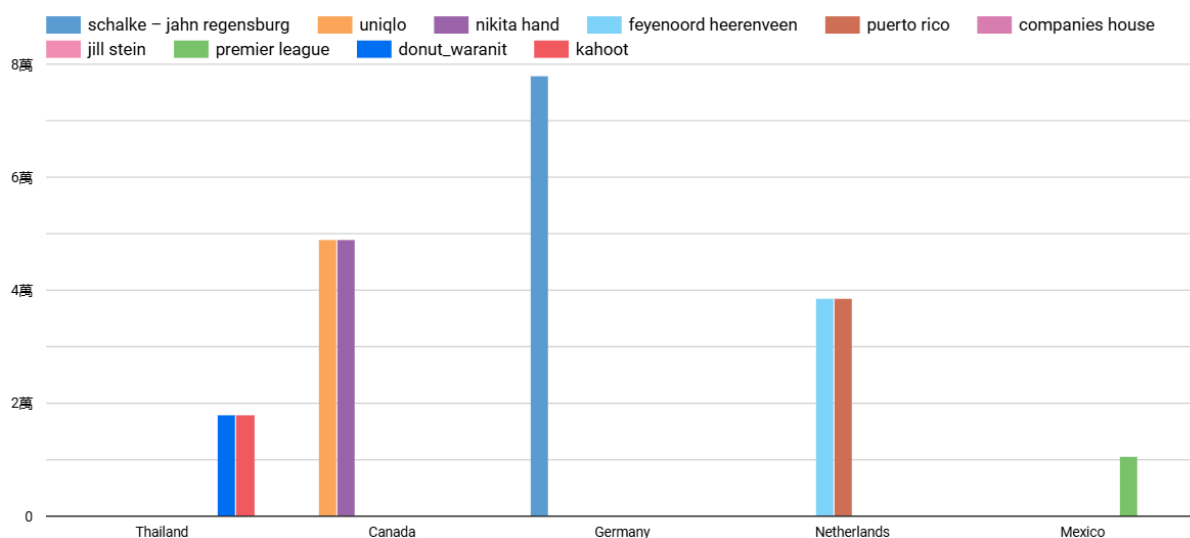
各表格的下鑽查詢皆有開啟

Conversion 和 CTR 的散佈圖



透過此圖，我們可以分析各廣告其點擊率和轉換率間的關係，因為顧客點擊廣告後，不一定能實質轉換，可能購買流程設計具有改善空間，或是顧客認為廣告與產品間有差距，無法達成購買。

分數最高的關鍵字之出現的次數



找出三個國家的business insights

巴西：在CTR的heatmap中，可以看出巴西的CTR相較於其他國家特別的突出。

這意味著巴西可能有以下的價值：

巴西的高CTR可能顯示出當地廣告的投放效果相對較好。這可能意味著廣告內容與當地受眾的需求和興趣匹配較高，或者是廣告展示的時機、位置選擇得當，從而吸引了更多的點擊。

巴西可能擁有較高的網絡使用率，這表明該國的數字廣告市場潛力巨大。對於企業來說，這意味著巴西是進一步加大廣告投放和擴展業務的重要市場。

巴西地區熱門的關鍵字皆與足球相關，若是能夠在足球場或足球轉播投放廣告，亦或是邀請明星足球員作為品牌代言人，搭配巴西地區傲視全球CTR，相信會有很好的成效。

越南：在google trends的資料集中，越南擁有最多數量的關鍵字。

我們認為可能有以下潛在的價值：

高用戶活躍度：越南網民在搜尋引擎上的參與度高，顯示市場潛力大，適合數位行銷。

需求多樣性：消費者需求多樣，企業可以根據不同興趣進行精準行銷。

增長潛力：越南可能正在快速發展，未來有潛在的消費升級機會。

數位化趨勢：網路普及率高，數位營銷和電子商務機會增多。

具體建議：而越南地區最熱門的幾項關鍵字，其內容大致都與彩券有關，若是要投放廣告，或是進行促銷活動，可考慮與彩券聯名。

印度：編號為AD020的廣告在印度地區，從CTR與conversion的散佈圖中可以發現，印度地區的CTR雖高，但是conversion較低。

這表明了廣告本身具有吸引力，但用戶未能如預期完成行動。可能原因有產品與廣告承諾不一致、結帳步驟過於複雜等原因。

建議：改善落地網頁的設計，確保內容與廣告信息一致，簡化流程，提高用戶信任感。深度了解印度客戶的需求，可結合熱門關鍵字進行分析。並進行測試或調整，例如使用A/B test，藉此提高轉換率。

6. Conclusion and Recommendations

結論

1. CTR與conversion

觀察各地區CTR與conversion之間的關係

CTR 較高，但轉換率中等：

說明用戶點擊後未必轉換，建議優化登錄頁或購買流程。

CTR 和轉換率均較高：

表明市場反應良好，應加大資源投入。

CTR 較低，轉換率更低：

可能是廣告內容或定位不佳導致。

2. 各地區熱門關鍵字

歐洲與南美洲

此地區最熱門的關鍵字大多與**足球**有關，可在足球場或足球轉播投放廣告，亦或是邀請明星足球員作為品牌代言人。

亞洲

最熱門的關鍵字常常與彩券或是賽馬等賽事相關，可往此方面發想廣告。

未來蒐集資料之建議

資料多樣性擴充：蒐集更多的資料，目前僅有17支廣告的數據而已，可加入更多廣告的資料進行分析，相信可以擁有更全面的了解。

資料時間跨度擴充：目前並非所有資料的時間區間皆相同，且有的資料集僅有一天的資料，而有些則跨度較長，可以補充目前資料缺失的時間維度，以進行更加深入的分析。

高階模型：使用更多目前最新的模型來學習資料集，做出更準確的預測。