

R語言與Python語言數據分析工程師養成班

- R/Python/Julia/SQL 程式設計與應用
(R/Python/Julia/SQL Programming and Application)
- 資料視覺化 (Data Visualization)
- 機器學習 (Machine Learning)
- 統計品管 (Statistical Quality Control)
- 最佳化 (Optimization)



李明昌博士

<https://www.youtube.com/@alan9956>

<http://rwepa.blogspot.com/>

alan9956@gmail.com

主題

- 1. 2024/12/26(四) 第1章 教務管理規定、數據分析概論 
- 2. 2025/01/02(四) 第2章 R語言簡介與基本運算功能
- 3. 2025/01/07(二) 第3章 各種物件資料的運算與處理
- 4. 2025/01/23(四) 第4章 數據的整合、清理與轉換
- 5. 2025/02/04(二) 第4章 數據的整合、清理與轉換【補充篇】MySQL資料庫應用

下載: https://github.com/rwepa/r_data_scientist/

第1章 教務管理規定、數據分析概論

- 1.1 教務管理規定
- 1.2 RWEPA簡介
- 1.3 數據分析概論
- 1.4 APC方法
- 1.5 數據分析暨視覺化應用案例
- 1.6 資料分析與視覺化工具介紹
- 1.7 R,Python專案報告

1.1 教務管理規定

教務管理規定

- 教務管理規定
- 課程大綱
- 課程評量
- 請假規範
- 結訓條件

參考上課說明

1.2 RWEPA簡介

RWEPA簡介: <http://rwepa.blogspot.com/>

- 姓名：李明昌 (ALAN LEE)
- 現職：中華R軟體學會 理事
臺灣資料科學與商業應用協會 理事
- 學歷：中原大學 工業與系統工程所 博士
- 經歷：
 - 淡江大學 兼任教師
 - 育達科技大學 兼任教師
 - 佛光大學 兼任教師
 - 國立台北商業大學 兼任教師
 - 東吳大學 兼任教師
 - 育達科技大學 資訊管理系(所) 專任助理教授
 - 崇友實業 行銷企劃專員
 - 國航船務代理股份有限公司 海運市場運籌管理員
- 大專院校、資策會、工業技術研究院、國家發展委員會、中央氣象局、公平交易委員會、衛生福利部、縣市政府與日本名古屋產業大學等公民營單位演講達354場，3284小時。
- 連絡資訊：alan9956@gmail.com



- iPAS 巨量資料分析師 證照推廣
- iPAS 營運智慧分析師 證照推廣

<https://www.youtube.com/@alan9956>

影片 ► 全部播放

以下为截图中显示的视频信息：

- RWEPA | shiny企業實務應用 第6集-小明算命師(下) -中文...
觀看次數：137次 • 1個月前
字幕
- RWEPA | 第5集- shiny小明算命師(上)【附小明算命師實...
觀看次數：143次 • 2個月前
字幕
- RWEPA | 第4集-shiny銷售儀表板【附小明算命師習題...
觀看次數：153次 • 3個月前
字幕
- RWEPA | shiny企業實務應用 第3集-佈置【附54萬筆銷售...
觀看次數：137次 • 3個月前
字幕
- RWEPA | shiny企業實務應用 第2集-下拉式選單【附習題...
觀看次數：130次 • 3個月前
字幕

以下为截图中显示的视频信息：

- RWEPA | R - shiny企業實務應用 第1集-白話shiny
觀看次數：212次 • 3個月前
字幕
- RWEPA | Power BI
觀看次數：45次 • 3個月前
字幕
- RWEPA | 程式設計(Python)
觀看次數：54次 • 3個月前
字幕
- RWEPA | Python - streamlit
(建立企業儀表板, 資料分析...
觀看次數：715次 • 4個月前
字幕
- RWEPA | R - 大量 Excel 處理
觀看次數：218次 • 4個月前
字幕

1.3 數據分析概論

數據分析

- 數據分析是一個涵蓋數據收集、清理、轉換、分析、視覺化與溝通等流程。
- 它涉及將原始數據轉換為有用的洞察力，以支援決策、問題解決與行動方案。
- 上面哪些步驟較困難?!
- 數據種類：百百種！



可擴展標記語言 (XML)

- XML (Extensible Markup Language) 是一種標籤語言，是一個簡單且非常彈性的文字資料格式
- XML 是從標準通用標記式語言 (SGML) 中簡化修改出來的
- XML 初期是用於大規模電子出版而設計的文件資料標準
- XML 並不是用來編排內容，而是用來描述資料。
- 它並沒有如同 HTML 一般的預設標籤，使用者需要定義描述資料的各種標籤
- XML 的官方 MIME 類型是 application/xml，副檔名是 .xml
- XML 已成為網頁及其他領域中，進行各種資料交換的重要技術標準
- 可使用 R, Python, Java, .NET 連結 XML

空氣品質指標(AQI)(歷史資料) - XML

- <https://data.gov.tw/dataset/151824>

```
aqx_p_488.xml □
1 <?xml version="1.0" encoding="UTF-8"?>
2 <aqx_p_488><data><sitename>屏東(枋山)</sitename><county>屏東縣</county><aqi>26
</aqi><pollutant></pollutant><status>良好</status><so2>0.1</so2><co>0.11</co><o3>32.8</o3><o3_8hr>
28.0</o3_8hr><pm10>5</pm10><pm2.5>1</pm2.5><no2>3.1</no2><nox>3.6</nox><no>0.4</no><windspeed>1
</windspeed><winddirec>85</winddirec><datacreationdate>2023-06-12 10:00
</datacreationdate><unit></unit><co_8hr>0.1</co_8hr><pm2.5_avg>1</pm2.5_avg><pm10_avg>4
</pm10_avg><so2_avg>0</so2_avg><longitude>120.651472</longitude><latitude>22.260899
</latitude><siteid>313</siteid></data><data><sitename>新北(樹林)</sitename><county>新北市
</county><aqi>38</aqi><pollutant></pollutant><status>良好</status><so2>0.3</so2><co>0.45</co><o3>
13.4</o3><o3_8hr>6.9</o3_8hr><pm10>18</pm10><pm2.5>6</pm2.5><no2>23.2</no2><nox>35.9</nox><no>12.6
</no><windspeed>0</windspeed><winddirec>152</winddirec><datacreationdate>2023-06-12 10:00
</datacreationdate><unit></unit><co_8hr>0.3</co_8hr><pm2.5_avg>7</pm2.5_avg><pm10_avg>24
</pm10_avg><so2_avg>0</so2_avg><longitude>121.38352778</longitude><latitude>24.94902778
</latitude><siteid>311</siteid></data><data><sitename>桃園(三民)</sitename><county>桃園市
</county><aqi>25</aqi><pollutant></pollutant><status>良好</status><so2>0.5</so2><co>0.44</co><o3>
22.2</o3><o3_8hr>15.2</o3_8hr><pm10>16</pm10><pm2.5>9</pm2.5><no2>15.5</no2><nox>25.2</nox><no>9.6
</no><windspeed>1</windspeed><winddirec>77</winddirec><datacreationdate>2023-06-12 10:00
</datacreationdate><unit></unit><co_8hr>0.5</co_8hr><pm2.5_avg>9</pm2.5_avg><pm10_avg>24
</pm10_avg><so2_avg>0</so2_avg><longitude>121.38352778</longitude><latitude>24.94902778
</latitude><siteid>311</siteid></data>
```

JSON

- JSON (JavaScript Object Notation)是一種輕量級的資料交換格式
- JSON 源自於JavaScript, JSON不是用於寫程式, 而是儲存與交換資料
- JSON 具有易於人們閱讀, 編寫與其他程式相容性高
 - 網頁應用: JavaScript, Java, Node.js, PHP, C#, R, Python, Julia等
 - 資料庫應用: NoSQL資料庫, 包括文件儲存格式, 例: MongoDB, CouchDB等
- JSON 的官方 MIME 類型是 application/json , 副檔名是 .json 。
- 建立 JSON 字串:
 - 陣列 (array) 可以用中括弧 [] 來寫入資料
 - 物件可 (object) 以用 大括弧{ } 來寫入資料

參考: <https://en.wikipedia.org/wiki/JSON>

JSON的資料型態

- 數值：十進位數，不能有前導0，可以為負數，可以有小數部分。還可以用e或者E表示指數部分。不能包含非數值，例如:不可使用NaN。不區分整數與浮點數。JavaScript用雙精度浮點數表示所有數值。
- 字串：以雙引號 "" 括起來的零個或多個Unicode碼位。支援反斜線開始的跳脫字元序列。
- 布林值：表示為**true**或者**false**。
- 值的有序列表 (array)：有序的零個或者多個值。每個值可以為任意類型。序列表使用**中括號[]**。元素之間用逗號分割。形如：[value, value]
- 物件 (object)：無序的「鍵-值對」(key-value pairs)，其中鍵以字串為主。建議但不強制要求物件中的鍵是獨一無二的。物件以**大括號{ }**表示。
例: {key1: value1, key2:value2, ...}。
- **null**類型：以 **null** 表示

參考: <https://en.wikipedia.org/wiki/JSON>

臺灣地區人工魚礁區範圍 - JSON

- <https://data.gov.tw/dataset/49099>

```
[  
  {  
    "field001": "10017",  
    "field002": "望海巷人工魚礁禁漁區",  
    "field003": "N25°08'36\"",E121°48'18\"",  
    "field004": "26M",  
    "field005": "88/12/27",  
    "field006": "半徑0.5浬範圍內",  
    "field007": "WGS-84"  
  },  
  {  
    "field001": "10017",  
    "field002": "大武崙人工魚礁禁漁區",  
    "field003": "N25°10'45\"",E121°42'40\"",  
    "field004": "20M",  
    "field005": "88/12/27",  
    "field006": "半徑0.5浬範圍內",  
    "field007": "WGS-84"  
  },  
]
```

哪裡找資料？

- <http://rwepa.blogspot.com/2019/09/dataset.html>
- <https://github.com/rwepa/DataDemo>

政府資料開放平台 <https://data.gov.tw/>

UCI Machine Learning Repository <https://archive.ics.uci.edu/datasets>

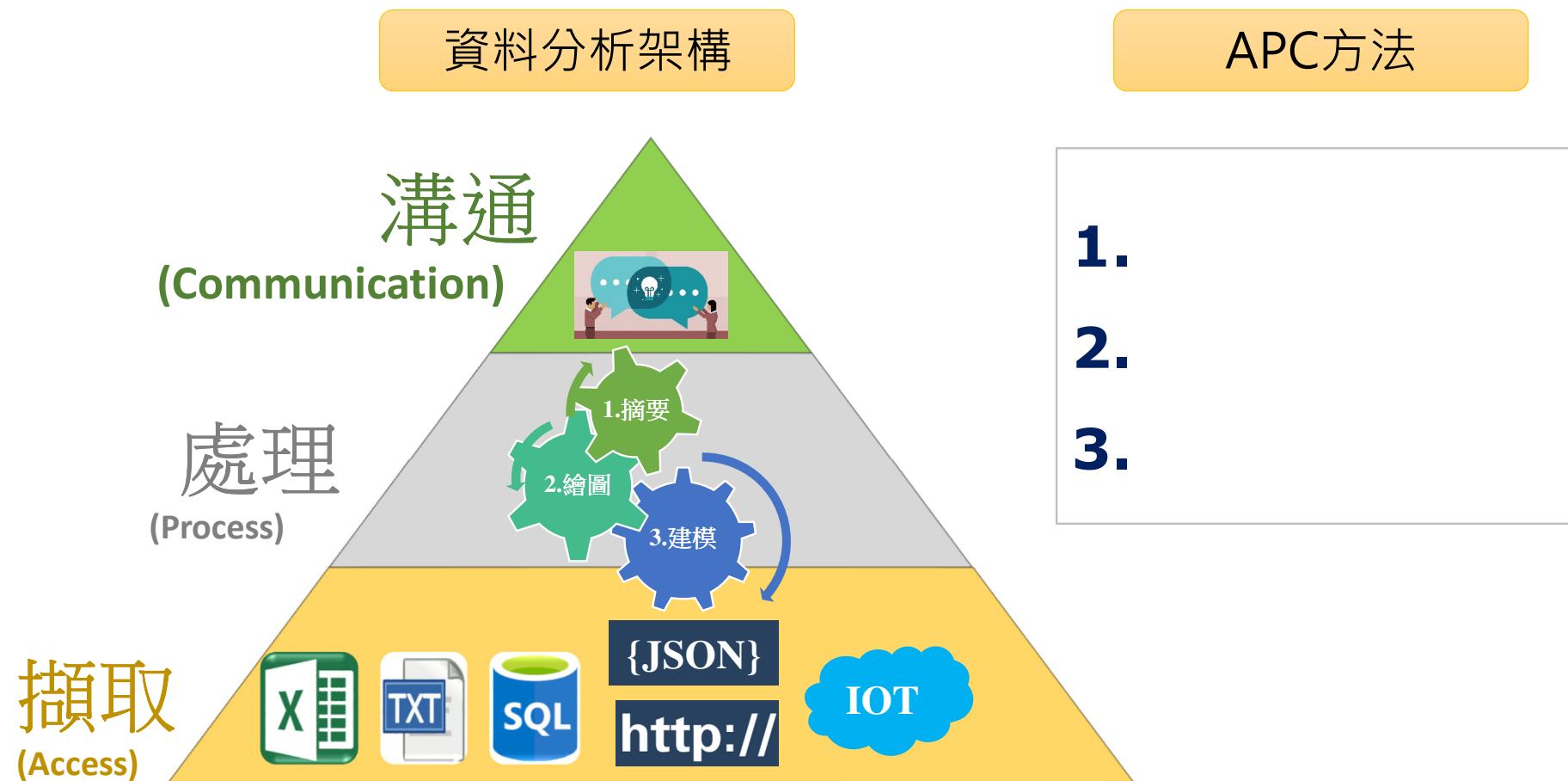
Google Dataset Search <https://datasetsearch.research.google.com/>

Kaggle Dataset <https://www.kaggle.com/datasets>

World Bank Open Data <https://data.worldbank.org/>

1.4 APC方法

★★★資料分析架構→APC方法



如何學習？

- 熟悉教材內容
- 將教材內容的資料集改為工作資料集(企業, 學術)
- 遇到問題時, 想辦法**尋找答案**
- 掌握 APC方法
- 掌握 ①摘要 ②繪圖 ③建模
- 參考網路應用文章 (進階) & 學術論文

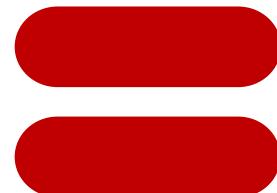
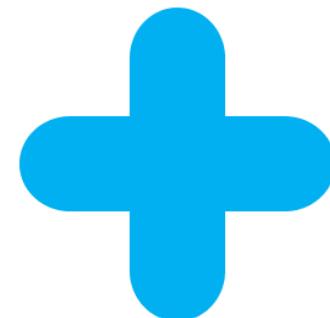
```
> 尋找答案 <- list(方法1 = c("同事", "同學", "朋友等"),
+                         方法2 = "Google",
+                         方法3 = "alan9956@gmail.com")
> print(尋找答案)
$方法1
[1] "同事"    "同學"    "朋友等"

$方法2
[1] "Google"

$方法3
[1] "alan9956@gmail.com"
```



學習目標



資料分析標準流程 - CRISP-DM

- 全名: 跨產業資料探勘標準流程
(CRoss Industry Standard Process for Data Mining)
- CRISP-DM 是屬於資料探勘方法論之一
- CRISP-DM是於1990年起，由SPSS以及NCR兩大廠商在合作戴姆克萊斯勒-賓士(Daimler Benz)的資料倉儲以及資料探勘過程中發展出來的。

CRISP-DM 資料探勘流程(續)

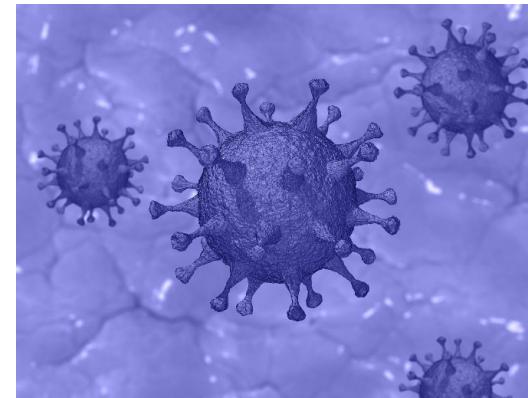
- 步驟 1：商業理解
- 步驟 2：資料理解
- 步驟 3：資料準備
- 步驟 4：模式建立
- 步驟 5：評估與測試
- 步驟 6：佈署應用

佔整專案時間的
~80%

- 訓練資料70%
- 測試資料30%

1.商業理解

- 終極目標是要解決具體的產業問題，諸如提高購買率、找出詐欺交易、銷售預測與異常偵測等，因此以專業知識 (domain knowledge)進行商業理解是重要的第一步，處理重點：
 - 擬定商業目標
 - 進行當前處境評估
 - 決定資料探勘目標：獲利、成本
 - 產生專案計劃
 - 解決顧客問題

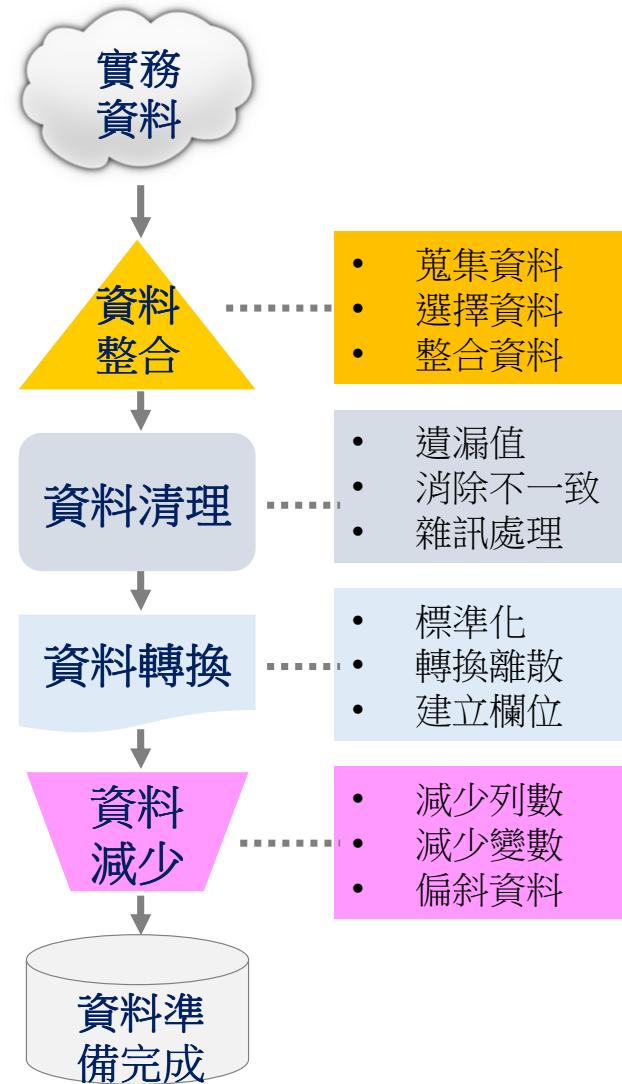


2. 資料理解

- 包括描述資料、探索資料、核驗資料品質
- 敘述統計分析
 - 六力分析(R: summary函數, Pandas: describe函數)
- 繪圖
 - 依群組特性
 - 依時間特性
 - 建立評估變數
 - 趨勢
 - 離群值 (outlier)
 - 散佈圖、散佈圖矩陣
 - 盒鬚圖
 - 樣式 (Patterns)

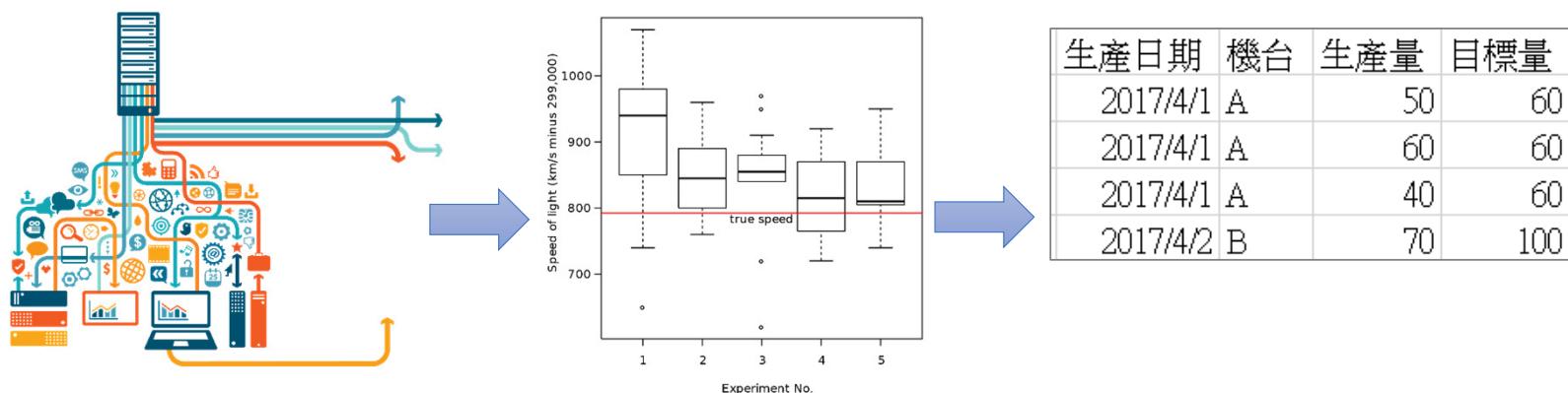
1. 最小值
2. 25百分位數 Q_1
3. 中位數 Q_2
4. 平均值
5. 75百分位數 Q_3
6. 最大值

3. 資料準備

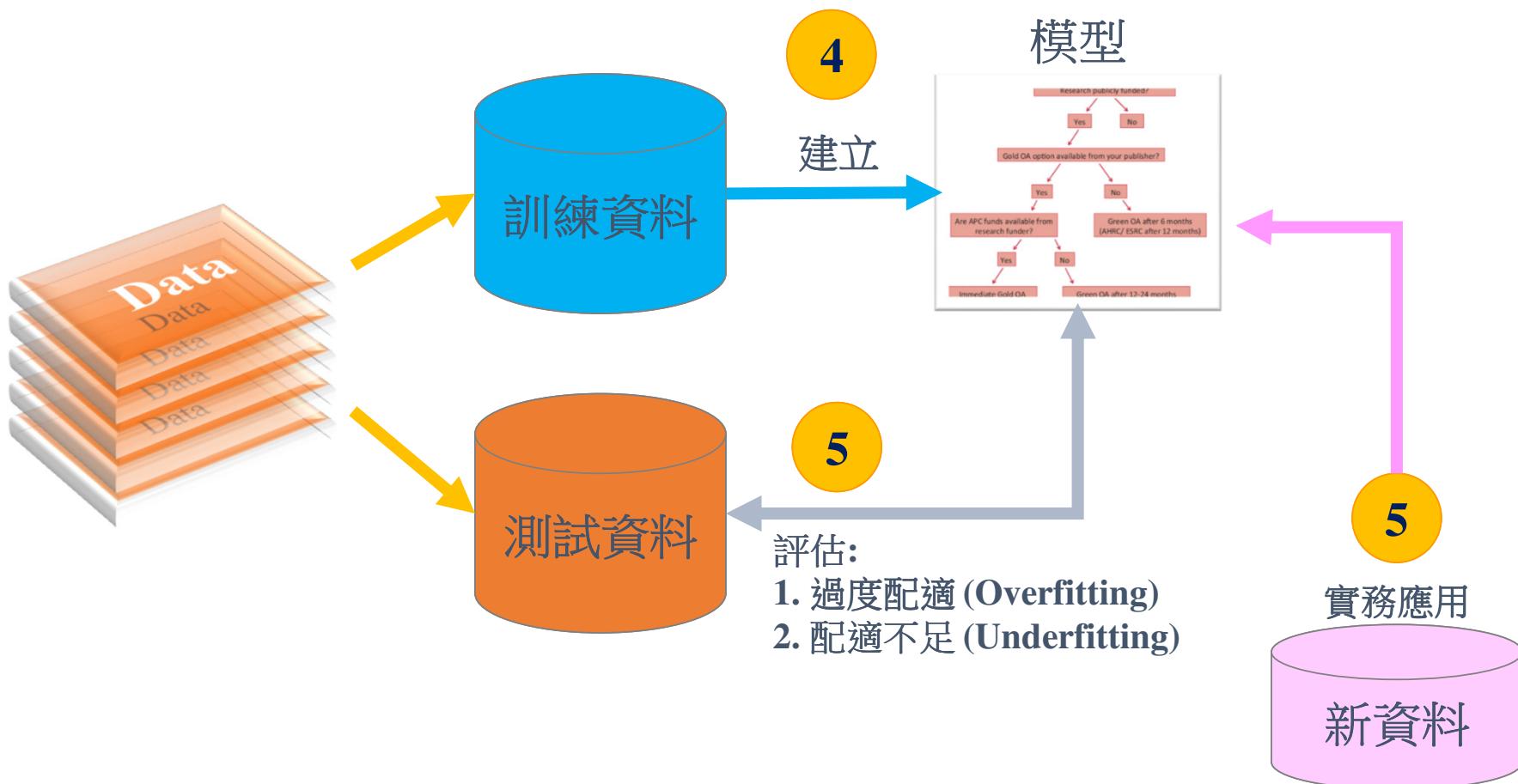


資料整合/理解/整理

- 資料整合: 從內部/外部獲得視覺化所需要的原始資料
- 資料理解: 了解資料之特性, 集中趨勢, 分散情形, 離群值 (Outliers)
- 資料整理: 對資料進行整理, NA值處理, 建立需求結構和格式



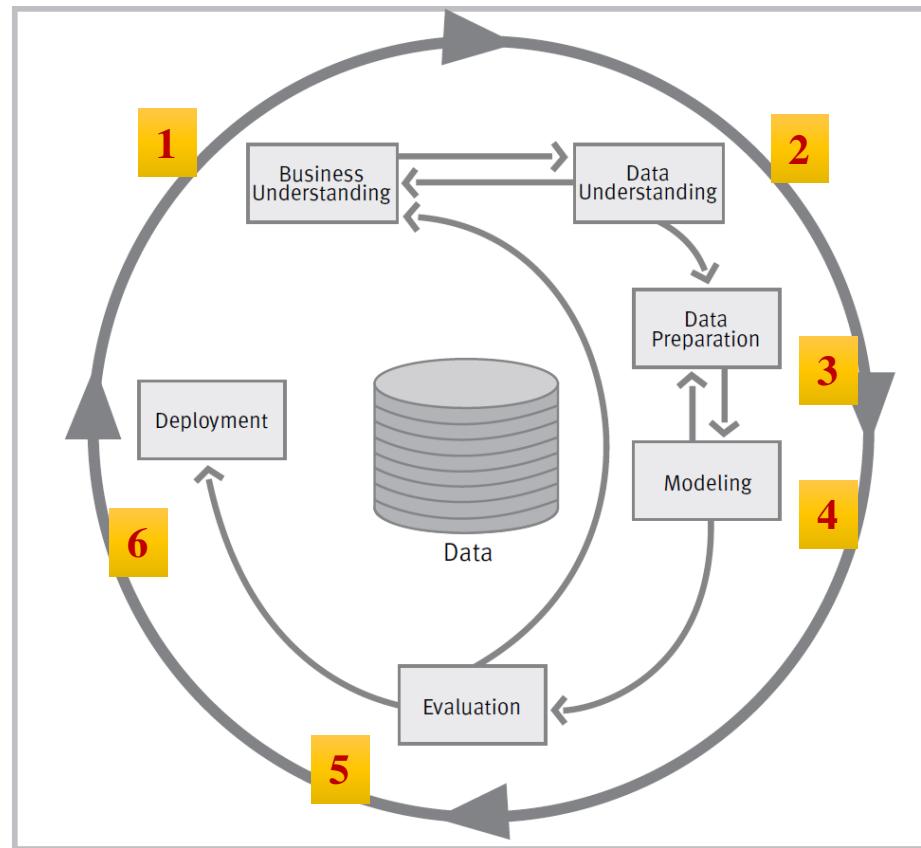
4.模型建立、5.評估與測試



6.佈署應用

- 資料分析專案的部署計畫(時間, 人力, 物力, 資源, \$)。
- 完成部署的服務應用，建立後續的監督管理與維運計畫(步驟)。
- 整個專案評估報告，包含商業獲利分析、資料分析結果、部署與維運管理計畫、成本效益分析、總結整體專案與下一階段計畫。
- CRISP-DM 運用步驟
 - 使用R/Python進行基礎資料分析與視覺化成果。
 - 建立機器學習預測模型(善用R的統計計算能力)。
 - 建立互動式資料分析(R: shiny套件, Python: streamlit模組)。
 - 善用Python深度學習應用。

CRISP-DM 資料探勘流程(續)



參考 https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

模型績效評估



數值模型績效指標

- 不可直接使用誤差的算術平均!

$$\text{Total error} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$$



- 均方誤差 (Mean Squared Error, MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- 均方根誤差 (Root Mean Squared Error, RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- 平均絕對誤差 (Mean Absolute Error, MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

類別模型績效指標

混淆矩陣
(Confusion Matrix)

- [RWEPA → rocr](#)
- <http://rwepa.blogspot.com/2013/01/rocr-roc-curve.html>

2020.3.23 中英文對照,Kappa統計量更新:

分類預測模型評估-混淆矩陣 (Confusion matrix)

混淆矩陣可用於監督式分類模型評估

	真實P類別	真實N類別
--	-------	-------

預測P類別 | TP真陽數 FP假陽數

預測N類別 | FN假陰數 TN真陰數

	P	N
--	---	---

1. TPR (True positive rate) 真陽性率, 愈大愈好
 $=TP / (TP+FN)$

$=TP/P$

=Sensitivity 瞩敏度

=Recall 召回率

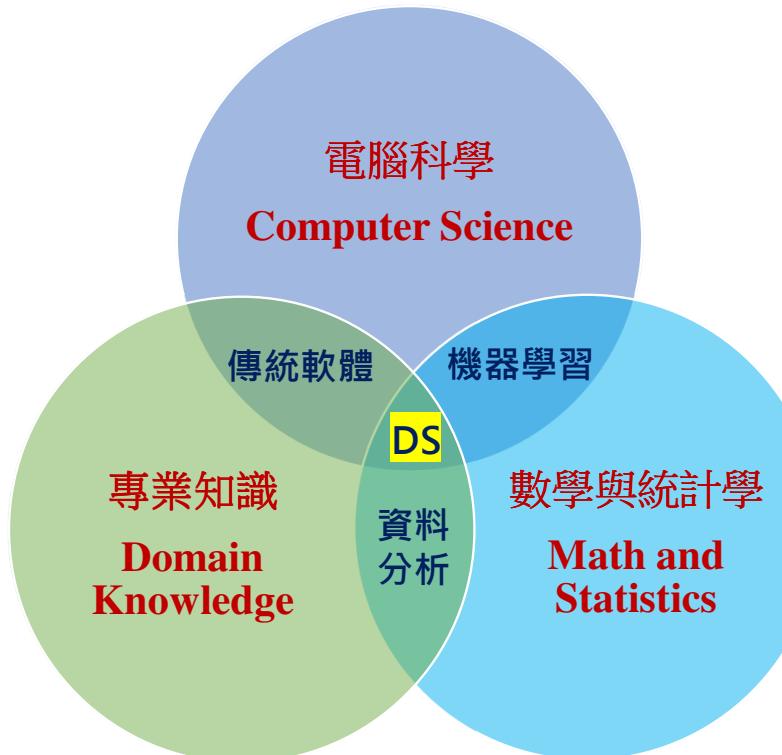
=Probability of detection

=Power

實際為陽性的樣本中，判斷為陽性的比例。

例如真正有生病的人中，被醫院判斷為有生病者的比例。

現代資料科學家 (Modern Data Scientist)



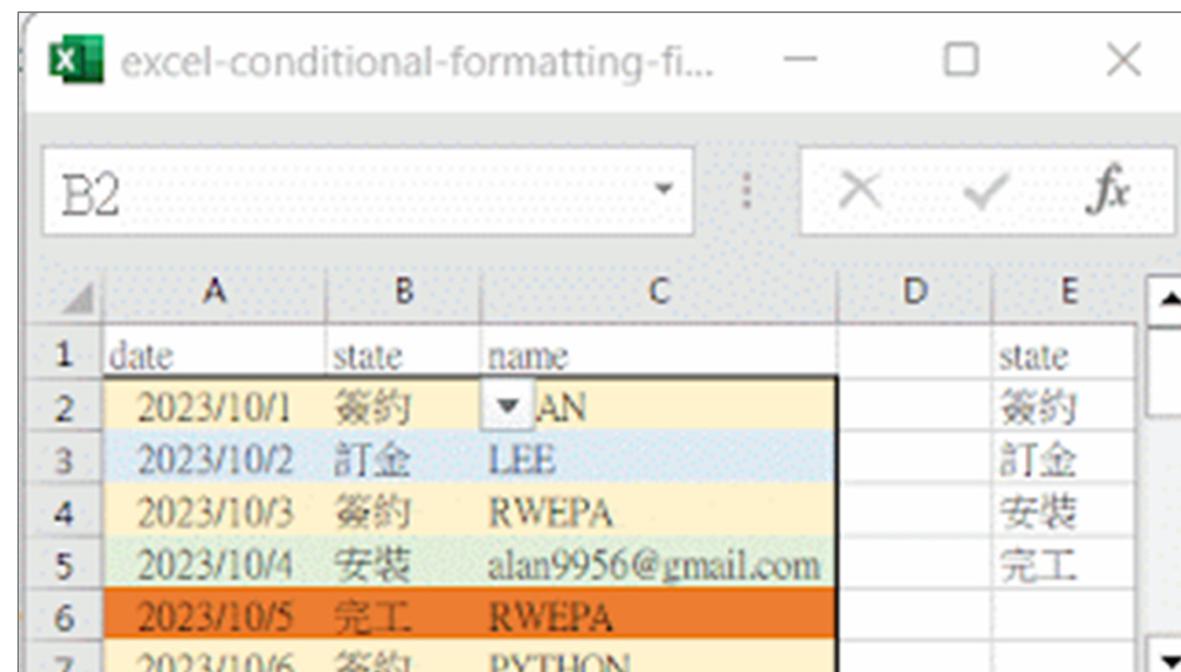
溝通與視覺化

- <https://towardsdatascience.com/a-long-term-data-science-roadmap-which-wont-help-you-become-an-expert-in-only-several-months-4436733e63ff>

1.5 數據分析暨視覺化應用案例

Excel 下拉式選單與條件式格式設定教學

- ✿ YouTube (包括中文字幕) : <https://youtu.be/OVA4dvkrsBM>
- LINK: <https://rwepa.blogspot.com/2023/10/excel-drop-down-list-and-conditional-formatting.html>

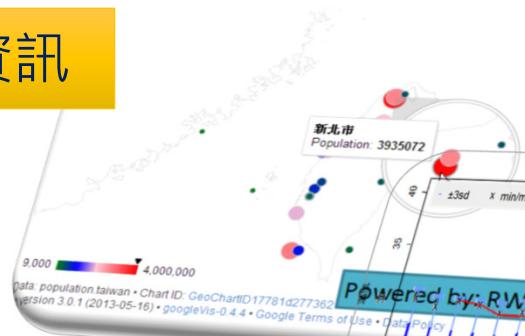


	A	B	C	D	E
1	date	state	name	state	
2	2023/10/1	簽約	AN	簽約	
3	2023/10/2	訂金	LEE	訂金	
4	2023/10/3	簽約	RWEPA	安裝	
5	2023/10/4	安裝	alan9956@gmail.com	完工	
6	2023/10/5	完工	RWEPA		
7	2023/10/6	簽約	PYTHON		

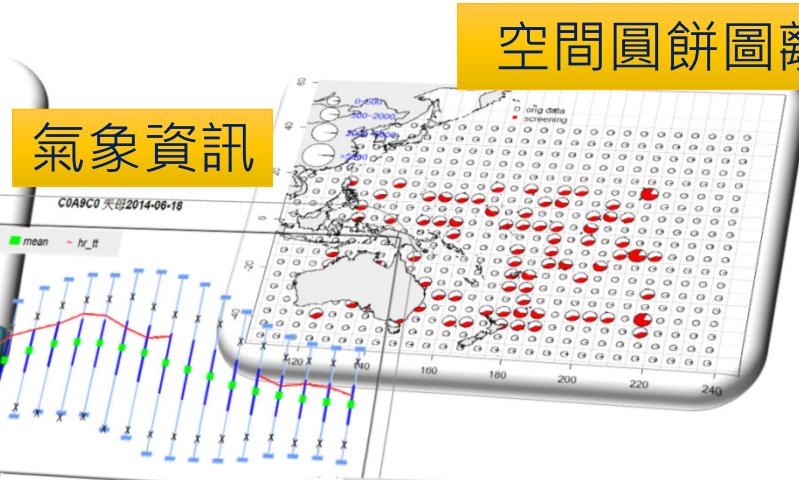
Excel 限制?

R+Shiny, Python+Streamlit 互動式平台

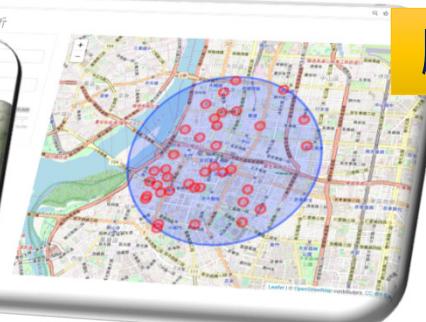
地理資訊



氣象資訊



空間圓餅圖離群值分析



保險預測



登山路線視覺化

顧客連結資訊



中央氣象局 1,600萬筆資料(14,328檔案)

網頁呈現

The screenshot shows a web-based interface for meteorological data analysis. On the left, there are several input fields for filtering data by station (466900 淡水), month (1-12), hour (1-24), year range (1998-2014), and QC threshold (-10 to 10). In the center, a scatter plot titled "QC_UKref/hr_tt_06mm_01hr_466900.txt" displays "實際值-TT" (Actual Value - TT) on the y-axis against "推估值-UK" (Predicted Value - UK) on the x-axis. A red 'X' marks a data point that appears to be an outlier. To the right, there's a histogram titled "學生化殘差直方圖" (Studentized Residual Histogram) showing the distribution of studentized residuals. At the top right, a summary box indicates "14,328 個檔案，18 個資料夾". The top navigation bar includes tabs for "一般" (General), "自訂" (Custom), and file icons.

異常偵測

A close-up view of the scatter plot from the central screenshot, highlighting a single data point marked with a red 'X' to indicate it as an outlier or abnormal value.

客製化選單

A screenshot of the platform showing a table of entries. The columns include Observation Station (466900), Date (20030601), Time (1), Actual Value TT (23.8), Predicted Value UK (20.6), Studentized Residual (5.163), CMT (23.751), SD (2.159), and Outliers (0.0227). The table shows one entry: "Showing 1 to 1 of 1 entries". Navigation buttons for "Previous" and "Next" are also visible.

動態繪圖

A histogram titled "學生化殘差直方圖" (Studentized Residual Histogram) showing the frequency distribution of studentized residuals. The x-axis ranges from -4 to 6, and the y-axis (Frequency) ranges from 0 to 200. The distribution is roughly symmetric and centered around zero.

R統計運算

A screenshot of R code output showing statistical results for the data entry. The results include the observation station (466900), date (20030601), time (1), actual value TT (23.8), predicted value UK (20.6), studentized residual (5.163), CMT (23.751), SD (2.159), and outliers (0.0227).

保險預測模型

機率模型閥值調整

The screenshot shows the iInsurance interactive analysis platform version v.16.3.24. The top navigation bar includes links for document upload, data processing, statistical charts, model evaluation, and prediction models. A red box highlights the 'Prediction Model' dropdown menu. Below it, a slider labeled 'Probability model threshold' is set to 0.1, with a red box around its value. A large yellow speech bubble on the left contains the text '機率模型閥值調整'. On the right, another yellow speech bubble contains the text '預測結果'. The main area displays a table of 12 entries with various columns including gender, vehicle type, exposure risk, and claim status. The last two columns, 'Predicted Probability' and 'Claim Status', are highlighted with red boxes. Red arrows point from the 'threshold' slider and the 'predicted probability' column to these highlighted areas.

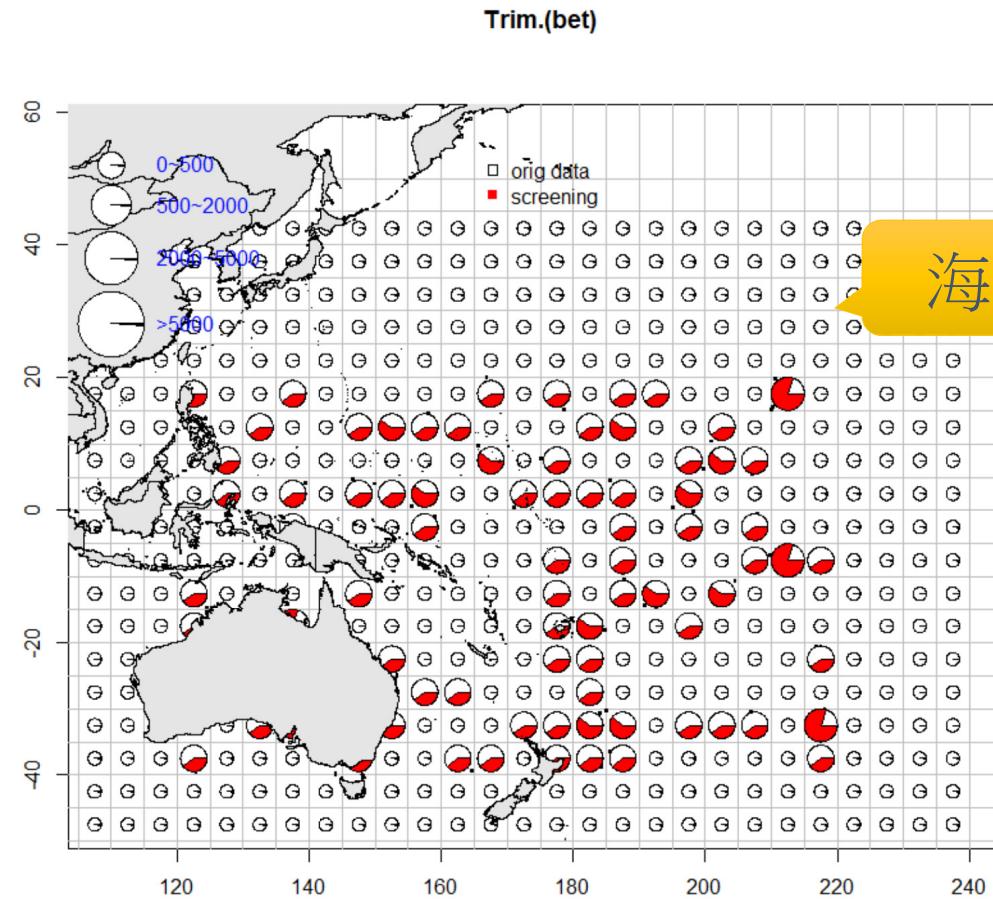
編號	性別	女性	車輛種類	私家車	曝露風險	曝露風險對數	無索償折扣	被保險人年齡	私家車 一車齡 0	私家車 一車齡 1	私家車 一車齡 2	私家車 車齡 0_1_2 組合	車齡 0_1_2 組合	預測機率	理賠	
1	M	0	A	1	0.9144422	-0.08944106	50	4	1	0	0	1	0	2	0.1069	有
2	M	0	A	1	0.8158795	-0.20348856	20	4	0	0	1	1	2	2	0.1441	有
3	M	0	A	1	0.8377823	-0.17699695	50	3	0	0	1	1	2	2	0.1866	有
4	M	0	A	1	0.4325804	-0.83798702	50	6	0	1	0	1	1	2	0.0944	無
5	M	0	A	1	0.7173169	-0.33223755	50	4	0	0	1	1	2	2	0.1218	有
6	M	0	A	1	0.8377823	-0.17699695	50	4	0	0	1	1	2	2	0.1495	有
7	M	0	A	1	0.8487337	-0.16400975	50	5	0	0	1	1	2	2	0.1422	有
8	F	1	A	1	0.8268309	-0.19015503	10	3	0	0	1	1	2	2	0.1733	有
9	M	0	A	1	0.7145791	-0.33606164	0	5	1	0	0	1	0	2	0.0694	無
10	M	0	A	1	0.3340178	-1.09656101	0	3	0	0	1	1	2	2	0.0783	無

Showing 1 to 10 of 12 entries

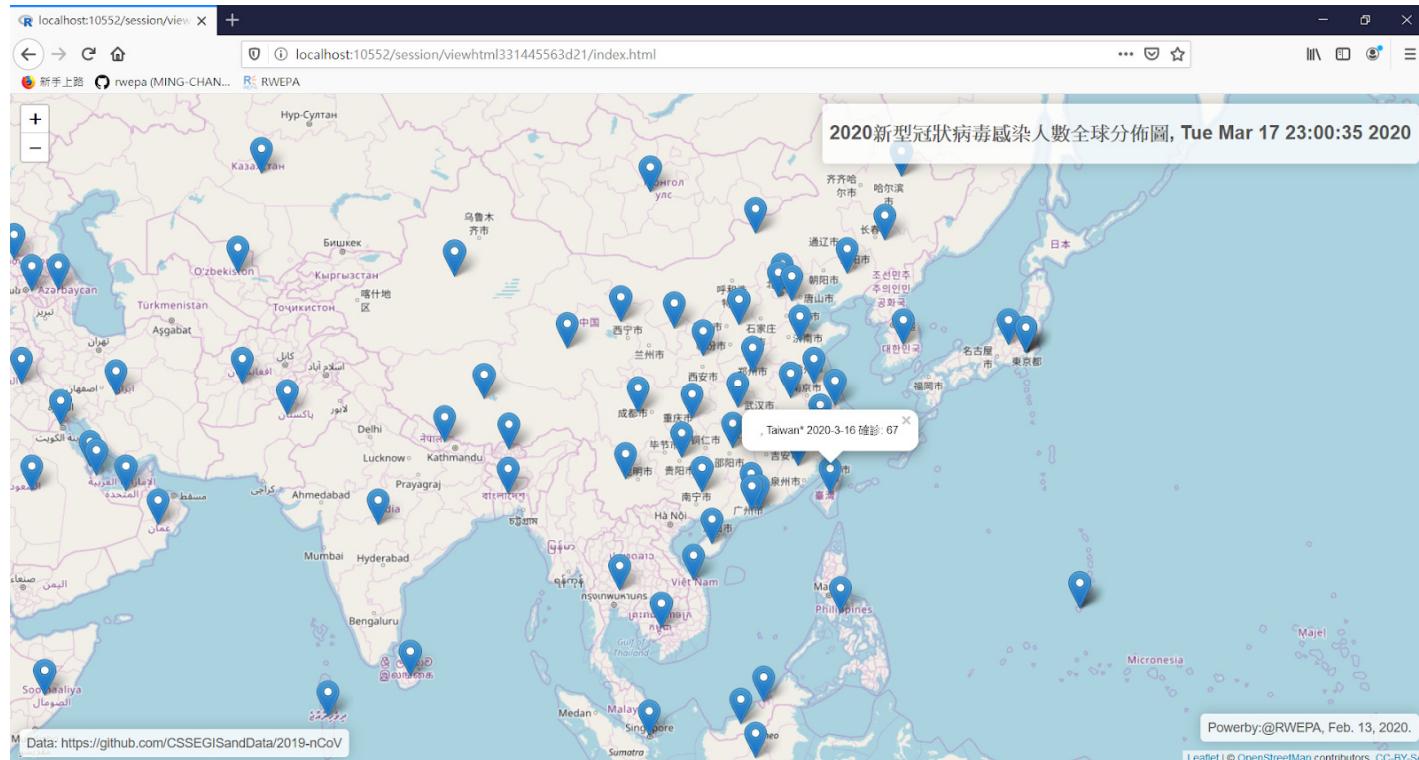
Previous 1 2 Next

127.0.0.1:6177/#tab-9487-2

空間圓餅圖離群值分析



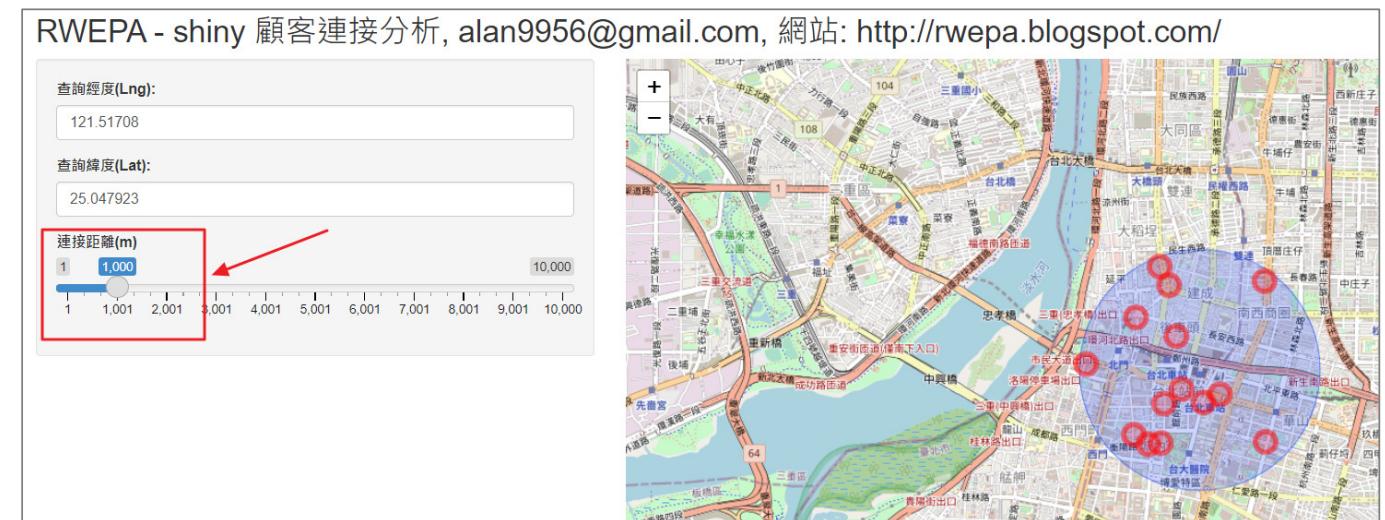
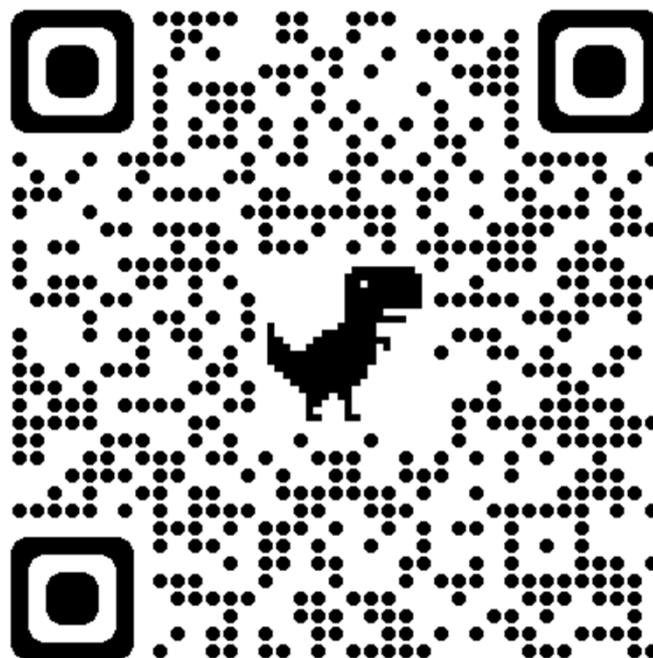
2020新型冠狀病毒視覺化



<http://rwepa.blogspot.com/2020/02/2019nCoV.html>

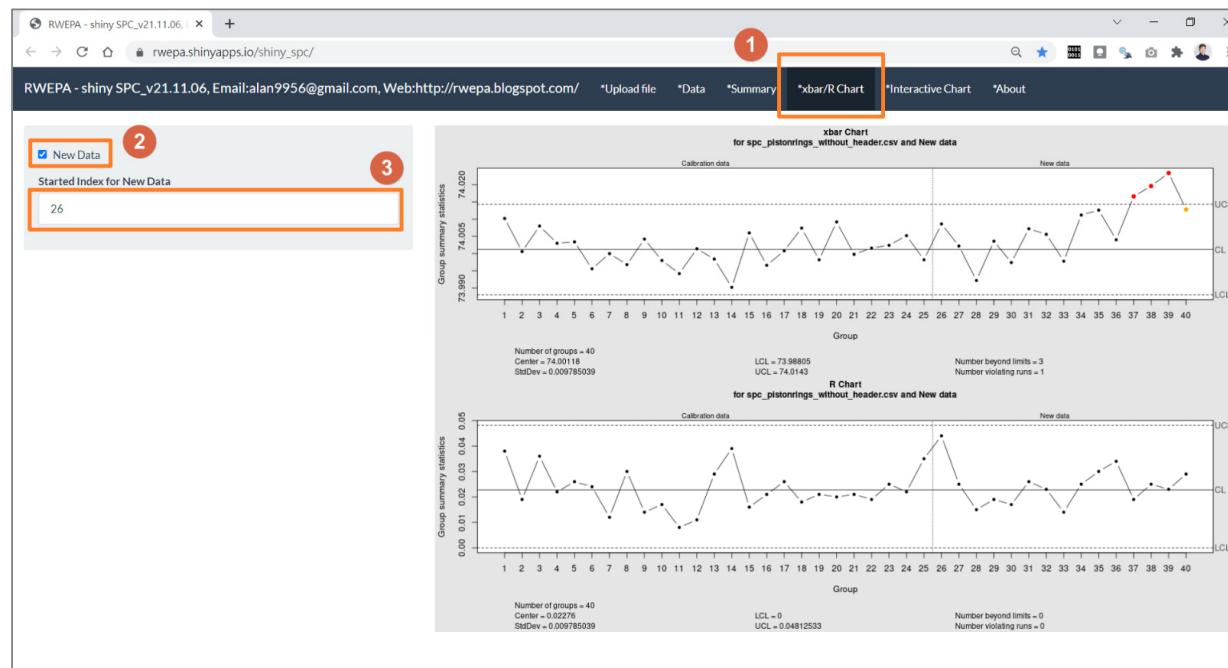
shiny 顧客連接分析

- <https://rwepa.shinyapps.io/shinyCustomerConnect/>



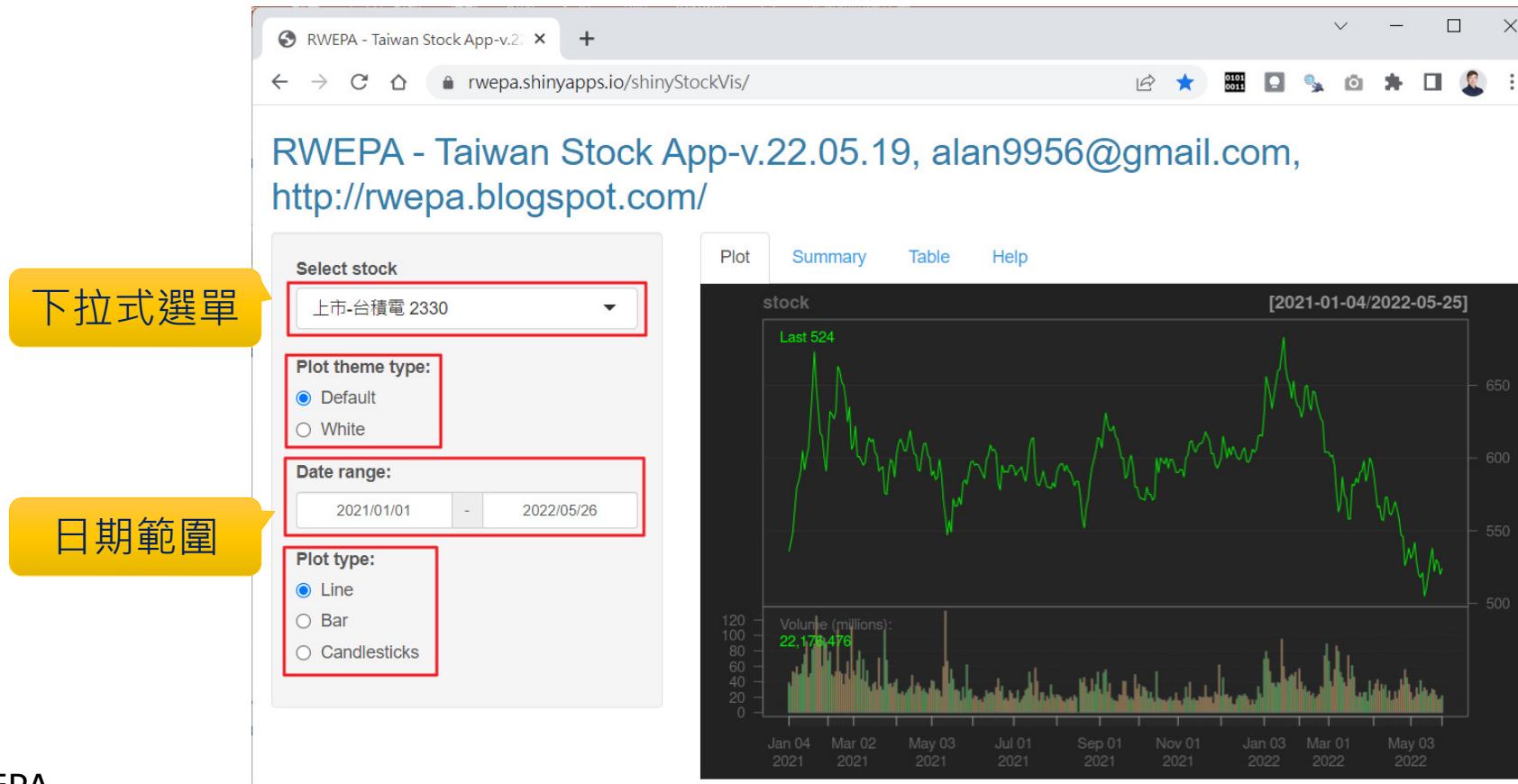
品質管制圖(quality control chart)應用

- 說明: <http://rwepa.blogspot.com/2021/10/r-shiny-quality-control-chart.html>
- 資料1: https://github.com/rwepa/shiny_spc/blob/main/data/spc_wafer_with_header.csv
- 資料2: https://github.com/rwepa/shiny_spc/blob/main/data/spc_pistonrings_without_header.csv
- 線上示範: https://rwepa.shinyapps.io/shiny_spc/

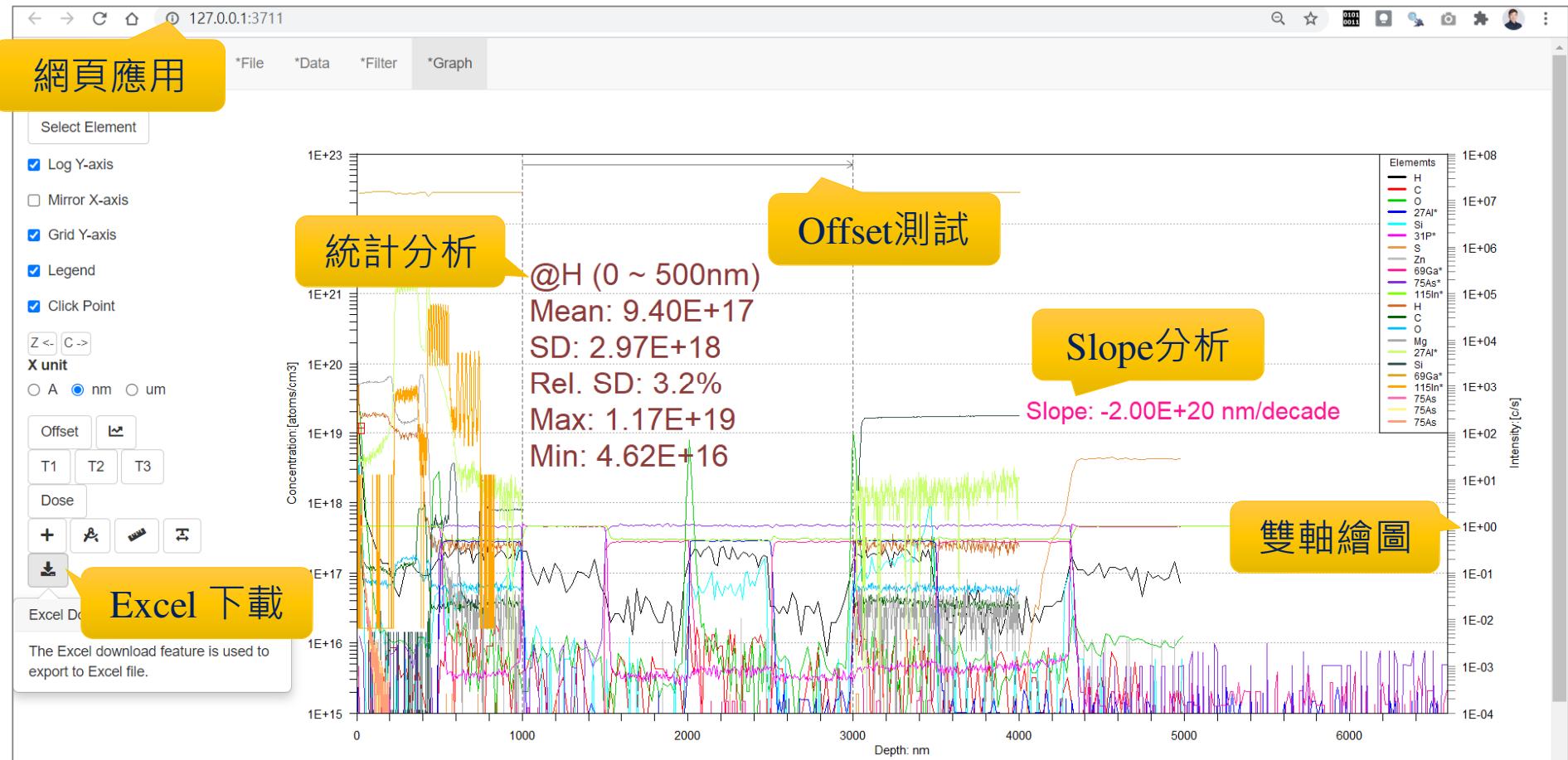


Taiwan Stock App

- <https://rwepa.shinyapps.io/shinyStockVis/>

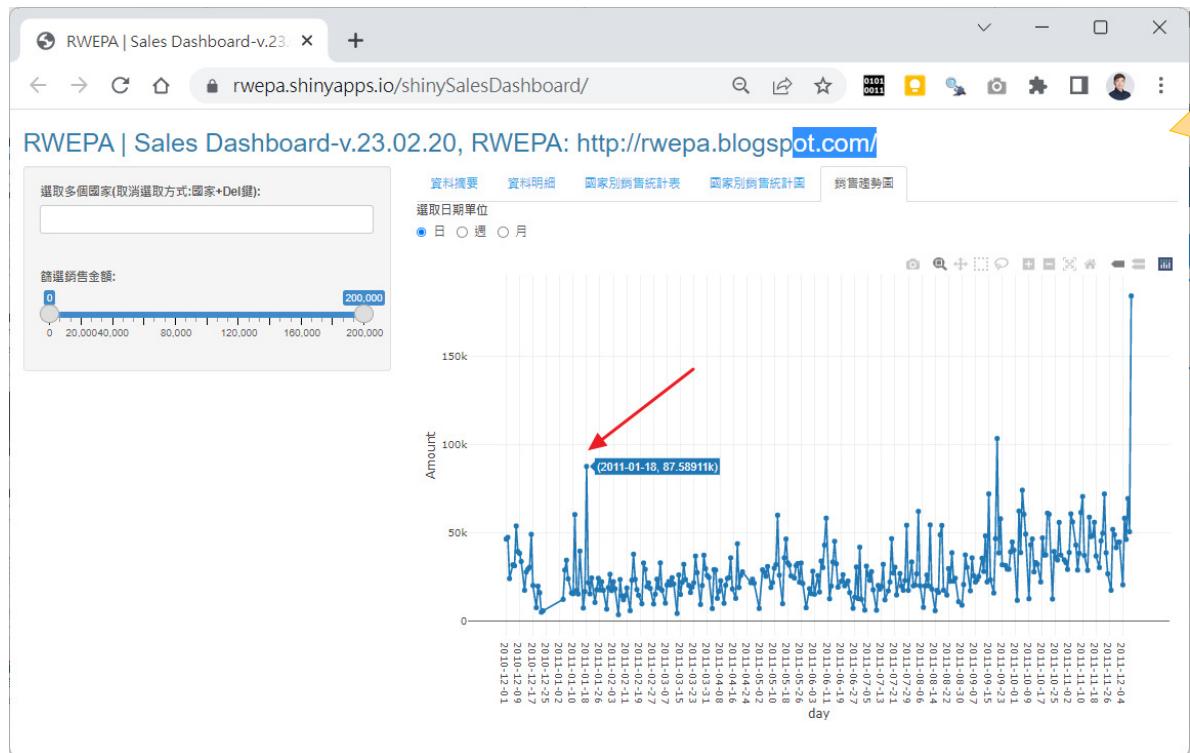


離子資料分析與視覺化應用



RWEPA | shiny企業實務應用 第4集-shiny銷售儀表板

- Shiny: <https://rwepa.shinyapps.io/shinySalesDashboard/>
- YouTube: <https://youtu.be/4GgZlf8heQk>



謝謝 ^_ ^

訂閱 + 讚 + 開啟小鈴鐺

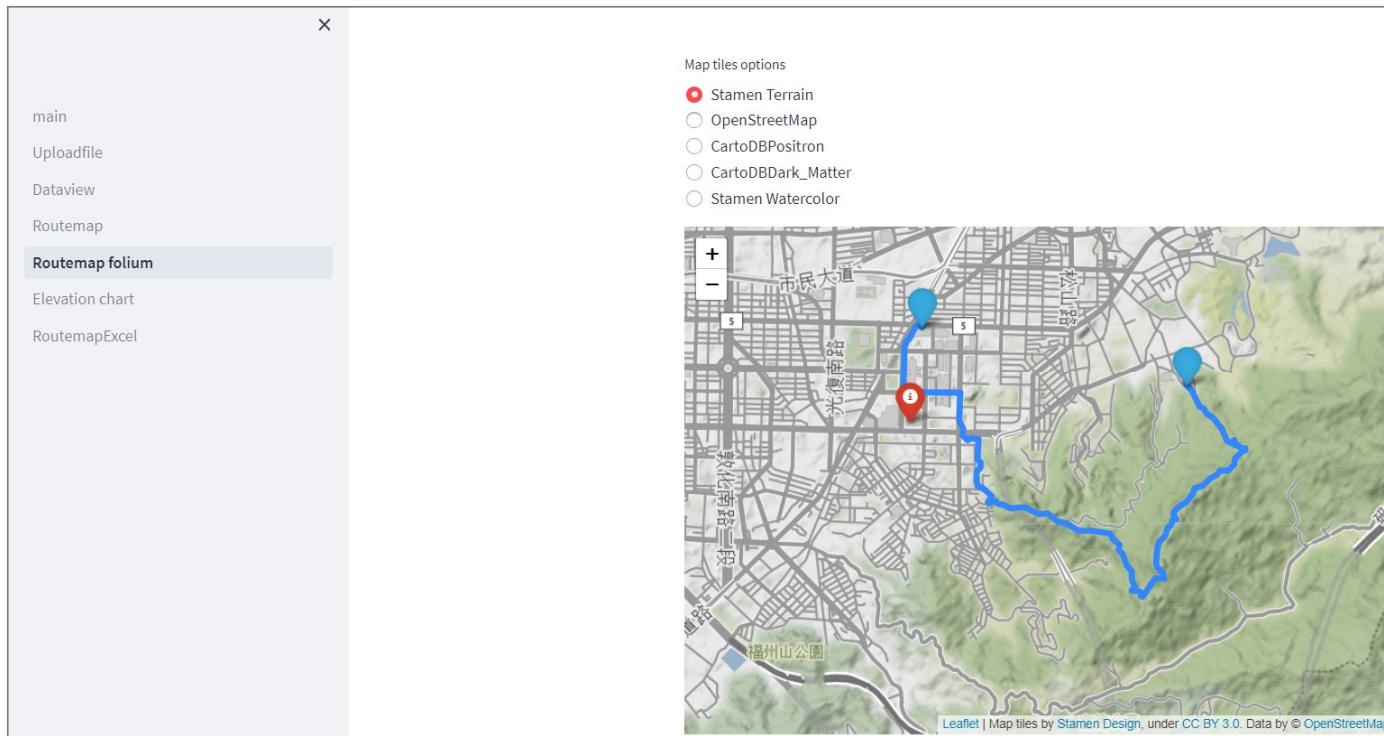
shiny企業實務應用 第6集-小明算命師(下) - 第1季完結篇

- Ubuntu Shiny Server: <https://shiny.rwepa.net/shiny-hr-teller/>
- YouTube: <https://youtu.be/rrD6KV3eV-w>



登山路線視覺化分析平台 (Python + Streamlit)

- YouTube : https://youtu.be/-_zghs2qrlg
- 系統展示 <https://rwepa-climb.streamlit.app/>



Power BI - 客戶輪廓分析

客戶輪廓Overview

客戶數: 666 | 平均年齡: 34.8

• FEMALE • MALE

288 (43.2%) 378 (56.8%)

客戶等級: 全選, VIP客戶, 一般客戶, 財富管理客戶

年齡組距: 01. 25歲以下 (11%), 02. 25-35歲 (48%), 03. 35-50歲 (32%), 04. 50-60歲 (5%), 05. 超過60歲 (3%)

職業: 其他未分類 (21%), 製造/營造業 (17%), 資訊科技業 (13%), 金融相關業 (11%), 餐飲服務業 (11%), 批發零售業 (10%), 軍/警/公/教 (9%), 醫療相關業 (5%), 運輸倉儲業 (2%), 專業人員 (1%), 服務業 (1%), 宗教、慈善 (1%), 政府官員 (1%), 教育業 (1%)

縣市別: 新北市 (29%), 台北市 (23%), 桃園市 (11%), 台中市 (8%), 高雄市 (8%), 台南市 (5%), 彰化縣 (4%), 基隆市 (3%), 新竹市 (2%), 新竹縣 (2%), 苗栗縣 (1%), 南投縣 (1%), 花蓮縣 (1%), 雲林縣 (1%)

區域別: 新店區 (4%), 內湖區 (4%), 中和區 (3%), 松山區 (3%), 大安區 (3%), 三重區 (3%), 中山區 (3%), 板橋區 (3%), 新莊區 (3%), 桃園區 (3%), 東區 (3%), 信義區 (3%), 中壢區 (2%), 中正區 (2%)

資料視覺效果:

- DAX量值管理表
- 分行資料表
- 日期對照表
- 刷卡交易資料表
- 刷卡國別對照表
- 刷卡類別對照表
- 客戶資料表
- 理財產品代碼表
- 理財產品交易資料表
- 業務員資料表
- Dynamic_Calendar

頁面 9 之 1

84%

Power BI – RFM分析

-  YouTube : <https://youtu.be/Lkr9HmzLTtg>
- LINK: <https://rwepa.blogspot.com/2023/07/rwepa-rfm-analysis-using-power-bi.html>

Customer Segmentation Using RFM Analysis, 2023



最近消費 (recency) :
顧客上次消費時間愈近，用戶價值愈大。

消費頻率 (frequency) :
顧客在一段時間中，總購買次數，購買頻率愈高，用戶價值愈大。

消費金額 (monetary) :
顧客總消費金額，消費金額愈高，用戶價值愈大。

Author : Ming-Chang Lee
YouTube : <https://www.youtube.com/@alan9956>
RWEPA : <http://rwepa.blogspot.tw/>
GitHub : <https://github.com/rwepa>
Email : alan9956@gmail.com

RFM分析 X RFM標準化分析 RECENTY FREQUENCY Monetary +

Tableau - Superstore

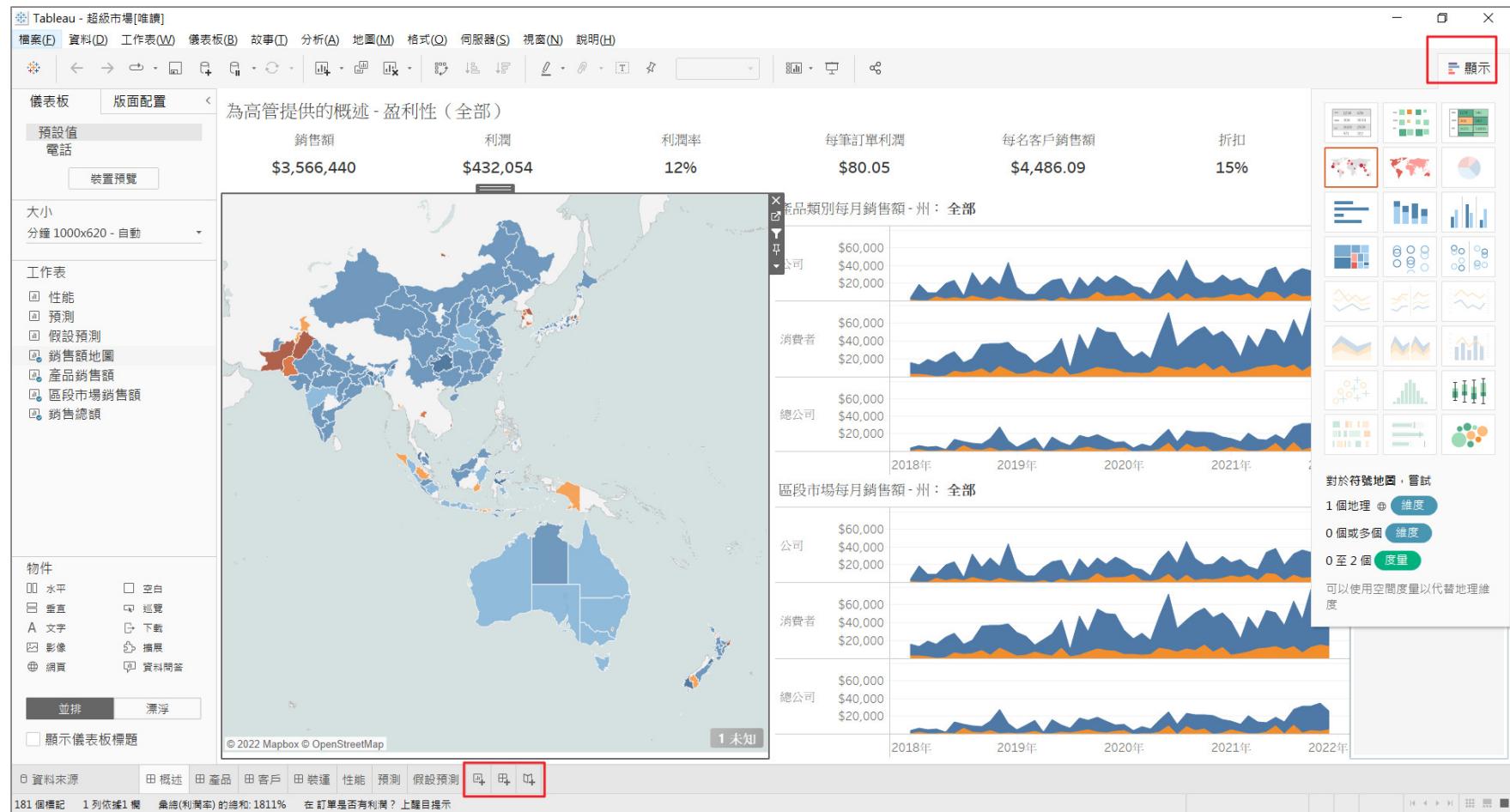
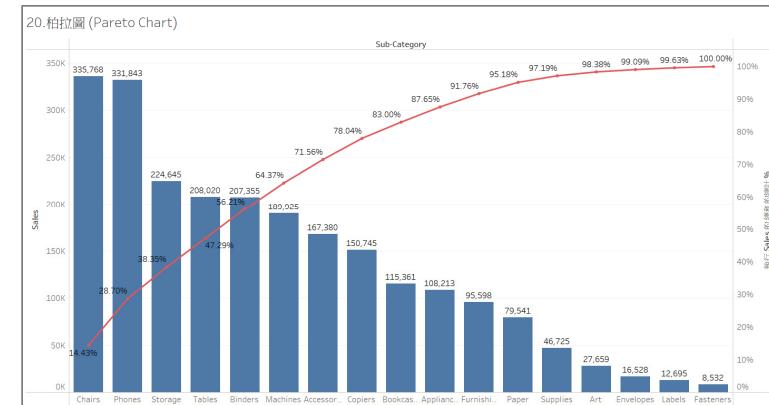
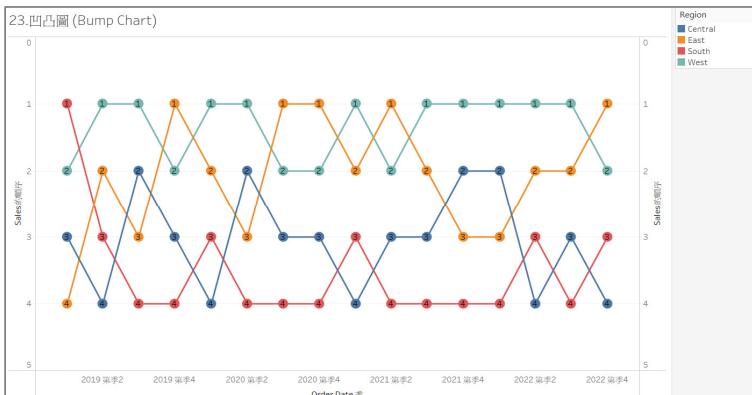
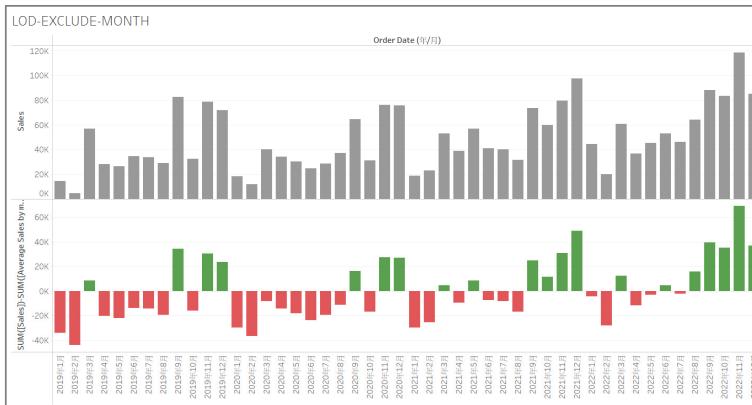


Tableau - 智慧製造應用

- <https://github.com/rwepa/Talks>
- <https://public.tableau.com/app/profile/ming.chang.lee/vizzes>



25.購物籃分析 (Market Basket Analysis)

Sub-Category

Sub-Cat.	Access.	Applianc...	Art	Binders	Bookcas...	Chairs	Copiers	Enviro...	Fasten...	Furnish...	Machin...	Paper	Phones	Storage	Supplies	Tables	
Accessories	718	60	89	161	29	64	7	32	27	115	47	22	153	118	103	25	44
Appliances	60	459	63	135	13	49	9	21	23	82	21	9	110	78	56	24	32
Art	89	63	756	159	34	87	7	29	30	105	47	12	152	124	100	29	41
Binders	161	135	159	1,339	56	126	19	54	62	200	82	30	276	199	201	43	74
Bookcases	29	13	34	56	228	21	2	11	7	30	21	6	48	42	37	10	9
Chairs	64	49	87	126	21	591	10	29	28	107	36	19	133	91	85	19	36
Copiers	6	9	7	19	2	10	70	4	5	12	6	2	20	11	14	6	4
Enviro...	32	21	29	54	11	29	4	251	11	32	11	6	59	41	38	5	12
Fasteners	27	23	30	62	7	28	5	11	226	48	15	8	59	39	39	13	14
Furnishings	115	82	105	200	30	107	12	32	48	919	57	29	181	154	140	31	46
Labels	47	21	47	82	21	36	6	11	15	57	348	8	80	58	61	18	17
Machines	22	9	12	30	6	19	2	6	8	29	8	114	28	22	21	1	6
Paper	153	110	152	276	48	133	20	59	59	181	80	28	1,205	179	178	45	54
Phones	116	78	124	199	42	91	11	41	39	154	58	22	179	826	117	22	54
Storage	103	56	100	201	37	95	14	38	39	140	61	21	178	117	797	33	50
Supplies	25	24	29	43	10	19	6	5	13	31	18	1	45	22	33	189	13
Tables	44	32	41	74	9	36	4	12	14	46	17	6	54	54	50	13	314

Tableau 教學

- Tableau資料分析與視覺化工具實作教師工作坊(初階)
 - https://github.com/rwepa/Talks/blob/main/tableau_tutorial_basic.pdf
- Tableau資料分析與視覺化工具實作教師工作坊(進階)
 - https://github.com/rwepa/Talks/blob/main/tableau_tutorial_advanced.pdf
- Tableau與R語言實務應用
 - https://github.com/rwepa/Talks/blob/main/tableau_r.pdf
- Tableau與MySQL資料庫實務應用
 - https://github.com/rwepa/Talks/blob/main/tableau_mysql.pdf



Python 程式設計-李明昌 免費電子書

- <http://rwepa.blogspot.com/2020/02/pythonprogramminglee.html>

主題: Python 程式設計-李明昌 - ipynb

檔名: Python_Programming_Lee_ipynb.zip

包括 Python 程式設計-李明昌電子書的原始 ipynb 檔案, 圖檔, 部分資料集

下載: https://github.com/rwepa/DataDemo/blob/master/Python_Programming_Lee_ipynb.zip



Python_Programming_Lee_ipynb.zip > python.book.lee >	
名稱	類型
.ipynb_checkpoints	檔案資料夾
data	檔案資料夾
img	檔案資料夾
Python程式設計-李明昌.ipynb	IPYNB 檔案

R 入門資料分析與視覺化應用(7小時28分鐘)

- <https://mastertalks.tw/products/r?ref=MCLEE>

課程提供教學範例的原始程式檔案與資料集 +中文字幕



- **主題**

1. R, RStudio簡介與套件使用
2. 認識資料物件
3. 資料處理與分析
4. 資料視覺化應用

- **特色**

1. 資料分析的**關鍵八步**
2. 提供必備**ggplot2**套件的應用知識與使用情境
3. 提供日期時間**zoo, xts**套件的整合應用操作
4. 提供**人力資源**資料與**銷售資料**，強化**實務資料**操作能力

R 商業預測應用(8小時53分鐘)

- <https://mastertalks.tw/products/r-2?ref=MCLEE>



- **主題**

1. R · RStudio工具操作
2. 非監督式學習商業預測
3. 監督式學習商業預測
4. 財金資料預測應用

- **特色**

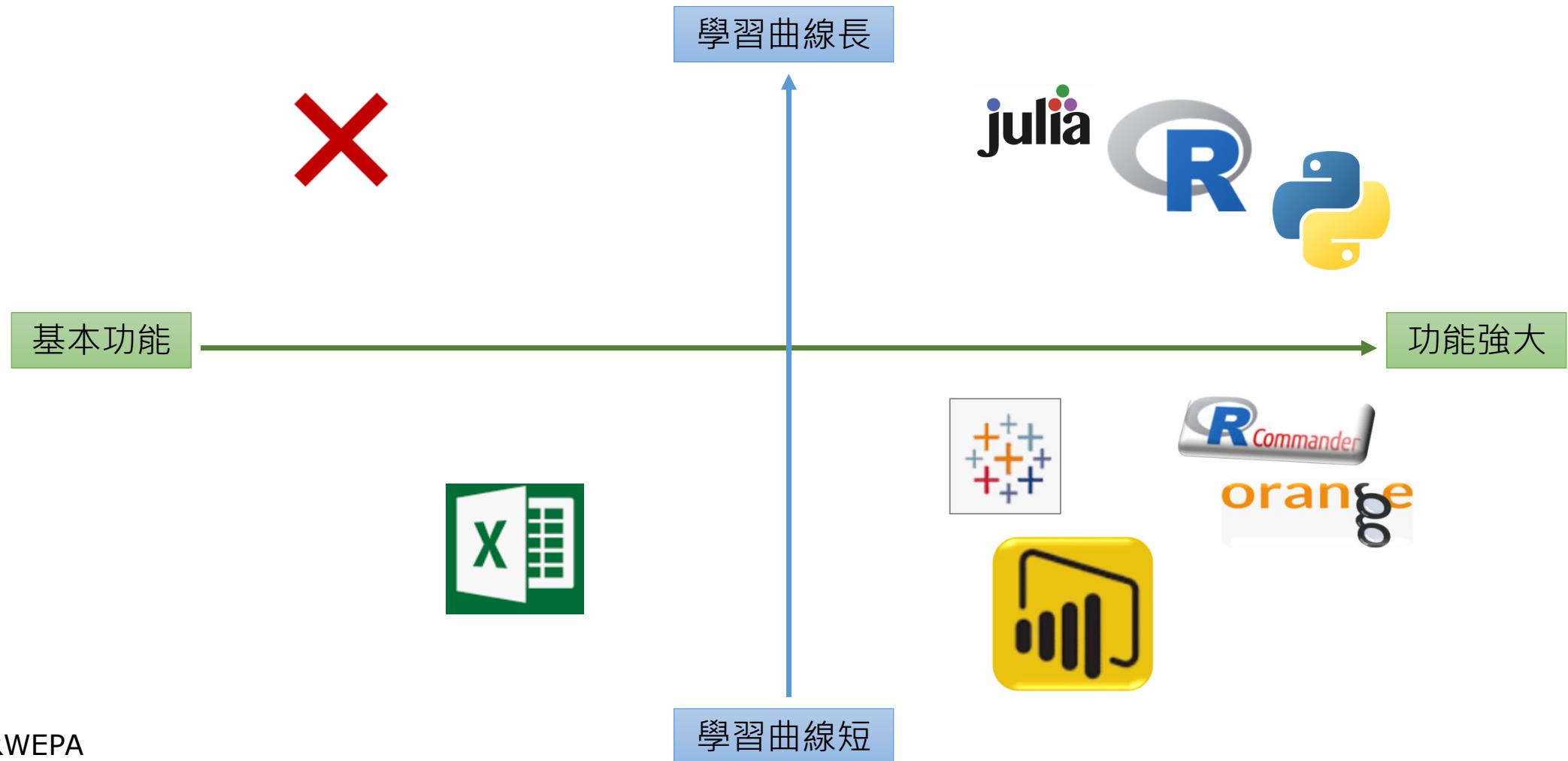
1. 採用**最有效率**方式學習大數據R語言，並應用於**職場資料分析**與**商業預測應用**
2. 提供**多元線性迴歸**的必備知識
3. 提供**財金資料商業預測應用**的基礎與進階必學技能
4. 提供學員人力資源資料與**台指期tick資料**預測演練

課程提供教學範例的原始程式檔案與資料集 +中文字幕

1.6 資料分析與視覺化工具介紹

- 1. Excel資料分析與視覺化
- 2. Power BI簡介
- 3. Tableau簡介
- 4. R簡介
- 5. Python簡介

資料分析與視覺化工具總覽



1. Excel資料分析與視覺化

(樞紐分析、合併彙算、迴歸預測)

Excel 條紋分析 - 線上交易銷售資料

- 資料網址：<https://www.kaggle.com/vijayuv/onlineretail>
- 檔名：OnlineRetail.csv.zip
- 解壓縮後檔名：OnlineRetail.csv
- 解壓縮後大小：43.4MB
- 下載：<https://github.com/rwepa/DataDemo/blob/master/OnlineRetail.csv.zip>

檢視資料

- 檔案筆數：541909列
- 欄位數：8欄

1. InvoiceNo	發票號碼
2. StockCode	條碼
3. Description	產品名稱
4. Quantity	數量
5. InvoiceDate	發票日期
6. UnitPrice	單價
7. CustomerID	顧客編號
8. Country	國家

A	B	C	D	E	F	G	H	
1	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
541905	581587	23256	CHILDRENS CUTLERY SPACEBOY	4	12/9/2011 12:50	4.15	12680	France
541906	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	12/9/2011 12:50	0.85	12680	France
541907	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	12/9/2011 12:50	2.1	12680	France
541908	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	12/9/2011 12:50	4.15	12680	France
541909	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	12/9/2011 12:50	4.15	12680	France
541910	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	12/9/2011 12:50	4.95	12680	France



找出有用的特徵(有意義的屬性)

- InvoiceNo
- Quantity
- InvoiceDate
- Unit Price
- Customer ID

- Q:企業最想分析哪些資料?
- A:每月收入統計圖

IMPORTANT



分割欄位 - 資料 \ 資料剖析

- InvoiceDate 右側 \ 插入1欄
- 選取InvoiceDate 資料
- 資料 \ 資料剖析
- InvoiceDate: 由1欄分割為2欄
- F2儲存格輸入 InvoiceTime



	A	B	C	D	E	F	G	H	I
1	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	2	UnitPrice	CustomerID	Country
2	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26		2.55	17850	United Kingdom
3	536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26		3.39	17850	United Kingdom
4	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26		2.75	17850	United Kingdom
5	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26		3.39	17850	United Kingdom
6	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26		3.39	17850	United Kingdom

InvoiceDate分割 3個欄位

- InvoiceDate分割 3個欄位:
InvoiceDateMonth, InvoiceDateDay, InvoiceDateYear

	C	D	E	F	G	H
1	Description	Quantity	InvoiceDateMonth	InvoiceDateDay	InvoiceDate Year	InvoiceTime
2	WHITE HANGING HEART T-LIGHT HOLDER	6	12	1	2010	08:26
3	WHITE METAL LANTERN	6	12	1	2010	08:26
4	CREAM CUPID HEARTS COAT HANGER	8	12	1	2010	08:26
5	KNITTED UNION FLAG HOT WATER BOTTLE	6	12	1	2010	08:26
6	RED WOOLLY HOTTIE WHITE HEART.	6	12	1	2010	08:26

建立新變數 InvoiceDateYearMonth, Revenue

- L2: =G2&TEXT(E2, "00") → 月份使用2位數值顯示

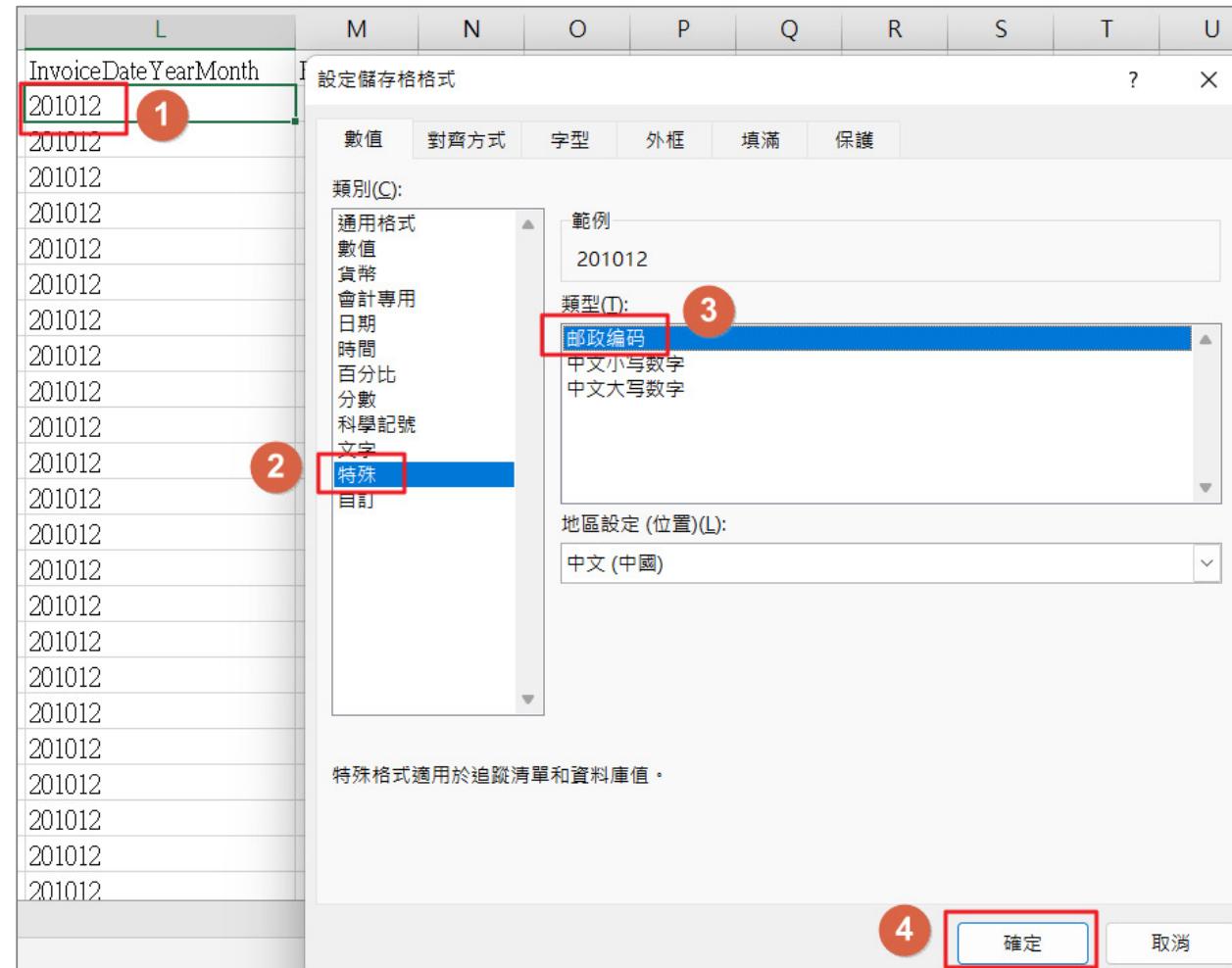
Excel screenshot showing the formula bar with the formula `=G2&TEXT(E2, "00")` highlighted in red. The formula is being entered into cell L2.

	E	F	G	H	I	J	K	L
1	InvoiceDateMonth	InvoiceDateDay	InvoiceDateYear	InvoiceTime	UnitPrice	CustomerID	Country	InvoiceDateYearMonth
2	12	1	2010	08:26	2.55	17850	United Kingdom	201012

Excel screenshot showing the formula bar with the formula `=D2*I2` highlighted in red. A red arrow points from cell D2 (Quantity) to cell I2 (UnitPrice). The formula is being entered into cell M2.

	D	E	F	G	H	I	J	K	L	M
1	Quantity	InvoiceDateMonth	InvoiceDateDay	InvoiceDateYear	InvoiceTime	UnitPrice	CustomerID	Country	InvoiceDateYearMonth	Revenue
2	6	12	01	2010	08:26	2.55	17850	United Kingdom	201012	=D2*I2

InvoiceDateYearMonth \ 格式 00000



新增 InvoiceDate

- N2: =DATEVALUE(G2 & "/" & E2 & "/" & F2)



The screenshot shows the 'Format Cells' dialog box from Microsoft Excel. The 'Number' tab is selected. In the 'Category' list, '短日期' (Short Date) is highlighted with a red border. The formula 'InvoiceDateValue' is displayed next to each category. On the right, a preview window shows four rows of data with the date '2010/12/1' repeated.

格式	公式
一般	InvoiceDateValue
數值	InvoiceDateValue
貨幣符號	InvoiceDateValue
會計專用	InvoiceDateValue
短日期	InvoiceDateValue
詳細日期	InvoiceDateValue
時間	InvoiceDateValue

N
InvoiceDate
2010/12/1
2010/12/1
2010/12/1
2010/12/1

樞紐分析表

	A	B
1		
2		
3	列標籤	<input type="checkbox"/> 加總 - Revenue
4	201012	748957.02
5	201101	560000.26
6	201102	498062.65
7	201103	683267.08
8	201104	493207.121
9	201105	723333.51
10	201106	691123.12
11	201107	681300.111
12	201108	682680.51
13	201109	1019687.622
14	201110	1070704.67
15	201111	1461756.25
16	201112	433668.01
17	總計	9747747.934

樞紐分析表欄位

選擇要新增到報表的欄位:

搜尋

InvoiceDateDay
 InvoiceDateYear
 InvoiceTime
 UnitPrice
 CustomerID
 Country
 InvoiceDateYearMonth
 Revenue

其他表格...

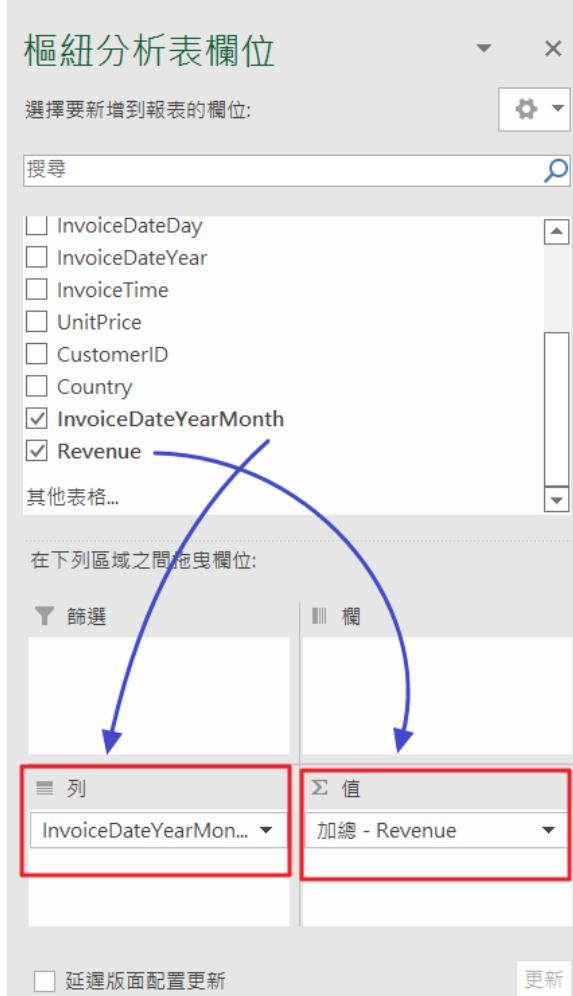
在下列區域之間拖曳欄位:

篩選 欄

列 值

InvoiceDateYearMon... 加總 - Revenue

延遲版面配置更新 更新



樞紐分析圖

- 1. 每月收入向上增加趨勢.
- 2. 2011年2,4月達到最低點.



範圍名稱

1 在工作表中選取範圍 A1:N9。

2 點選 Excel 功能選單上的「公式」。

3 點選「定義名稱」下的「從選取範圍建立」。

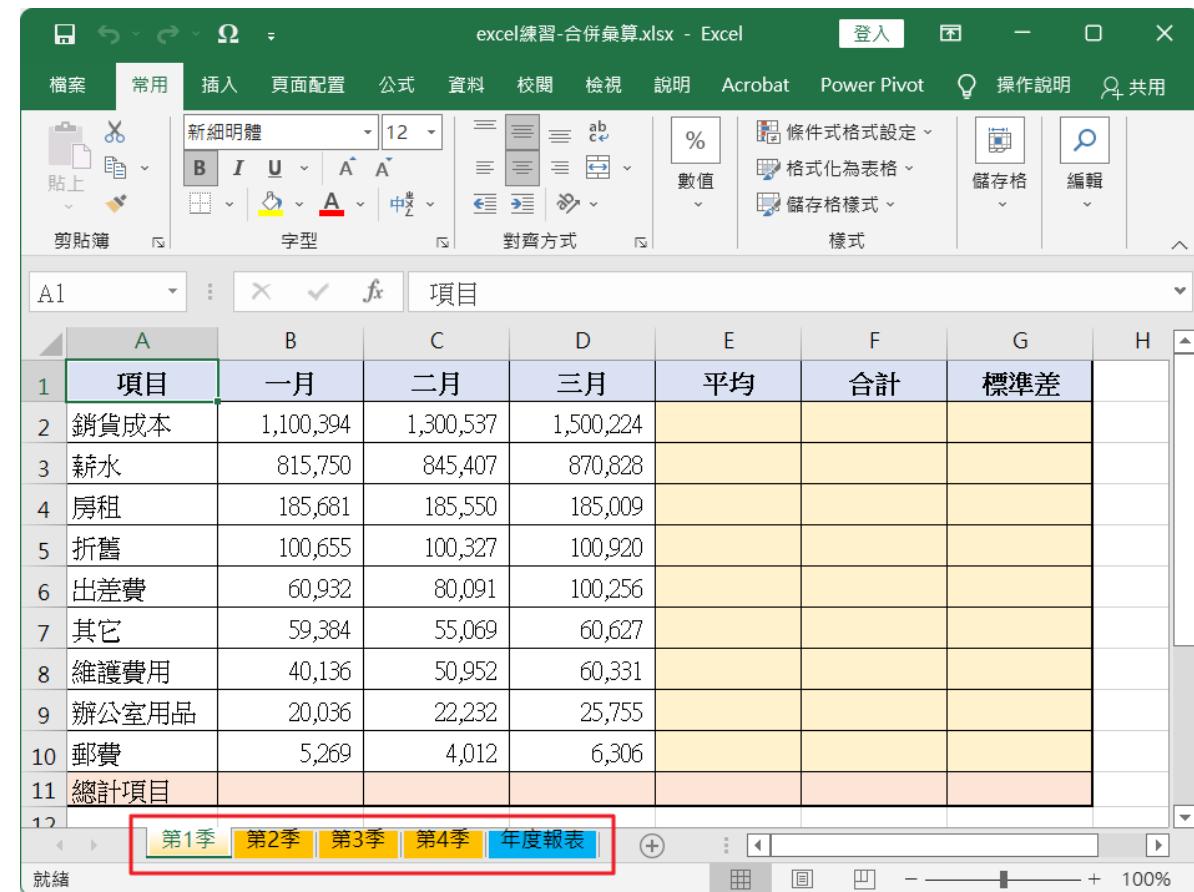
4 在「以選取範圍建立名...」對話框中，點選「頂端列(I)」並點選「確定」。

5 在「以選取範圍建立名...」對話框中，點選「確定」。

	G	H	I	J	K	L	M	N
1	InvoiceDateYear	InvoiceTime	UnitPrice	CustomerID	Country	InvoiceDateYearMonth	Revenue	InvoiceDate
1	2010	08:26	2.55	17850	United Kingdom	201012	15.3	2010/12/1
1	2010	08:26	3.39	17850	United Kingdom	201012	20.34	2010/12/1
1	2010	08:26	2.75	17850	United Kingdom	201012	22	2010/12/1
1	2010	08:26	3.39	17850	United Kingdom	201012	20.34	2010/12/1
1	2010	08:26	3.39	17850	United Kingdom	201012	20.34	2010/12/1
1	2010	08:26	7.65	17850	United Kingdom	201012	15.3	2010/12/1
1	2010	08:26	4.25	17850	United Kingdom	201012	25.5	2010/12/1
1	2010	08:28	1.85	17850	United Kingdom	201012	11.1	2010/12/1

合併彙算

- 平均值 =AVERAGE
- 合計 =SUM
- 標準差(?)
 - =STDEV.P()
 - =STDEV.S()

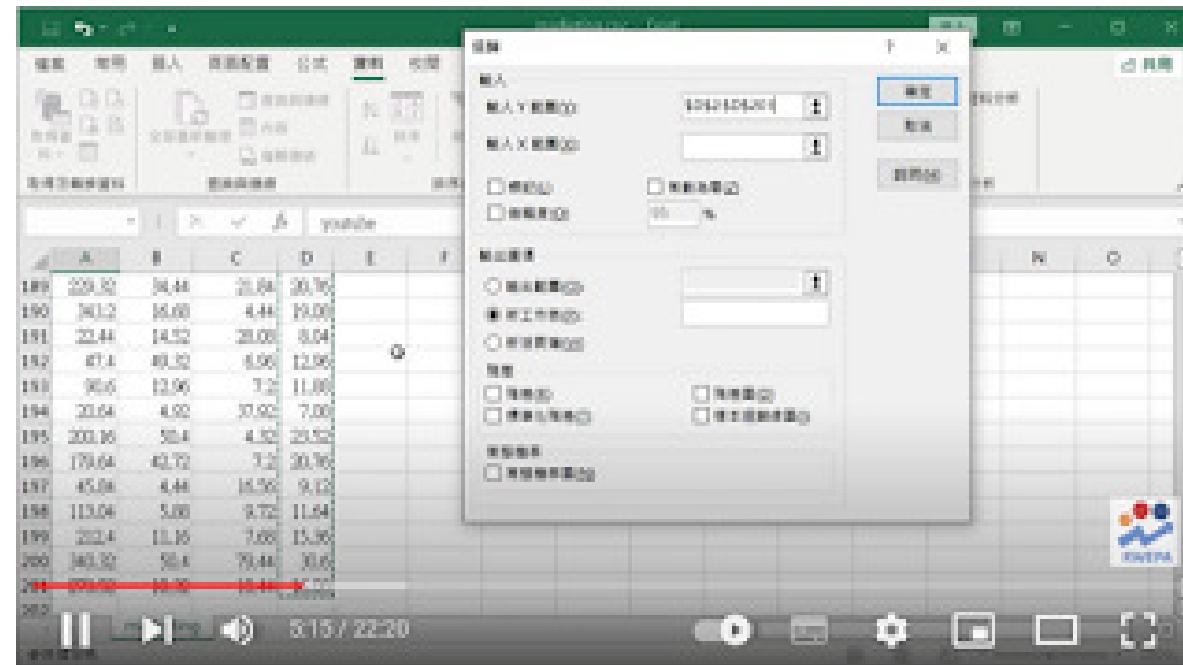


The screenshot shows an Excel spreadsheet titled "excel練習-合併彙算.xlsx". The table has columns for Project (項目), January (一月), February (二月), March (三月), Average (平均), Total (合計), and Standard Deviation (標準差). The last row is a summary for "Total Project" (總計項目). The bottom navigation bar shows tabs for "Q1", "Q2", "Q3", "Q4", and "Annual Report" (年度報表), with "Q4" and "Annual Report" highlighted in blue.

項目	一月	二月	三月	平均	合計	標準差
銷貨成本	1,100,394	1,300,537	1,500,224			
薪水	815,750	845,407	870,828			
房租	185,681	185,550	185,009			
折舊	100,655	100,327	100,920			
出差費	60,932	80,091	100,256			
其它	59,384	55,069	60,627			
維護費用	40,136	50,952	60,331			
辦公室用品	20,036	22,232	25,755			
郵費	5,269	4,012	6,306			
總計項目						
12	第1季	第2季	第3季	第4季	年度報表	

Excel 回歸分析

- YouTube: https://www.youtube.com/watch?v=i5_urp8XzEs
- 說明: <http://rwepa.blogspot.com/2022/05/httpsrwepa.blogspot.com202205regression-analysis-in-excel.html.html>
- 資料檔: <https://github.com/rwepa/DataDemo/blob/master/marketing.csv>



Excel 圖表

- 包括直條圖 ... 組合圖
- 合計17種圖表



2. Power BI簡介

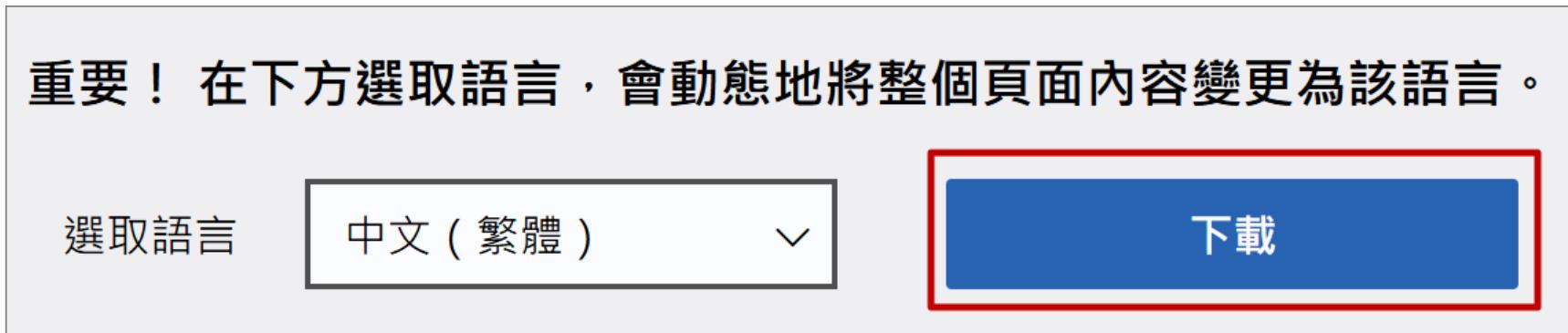
Power BI 下載

- <https://powerbi.microsoft.com/zh-tw/desktop/>
→進階下載選項



Power BI Desktop

- <https://www.microsoft.com/zh-TW/download/details.aspx?id=58494>
→ 下載



下載 PBIDesktopSetup_x64.exe

選擇您要的下載 X

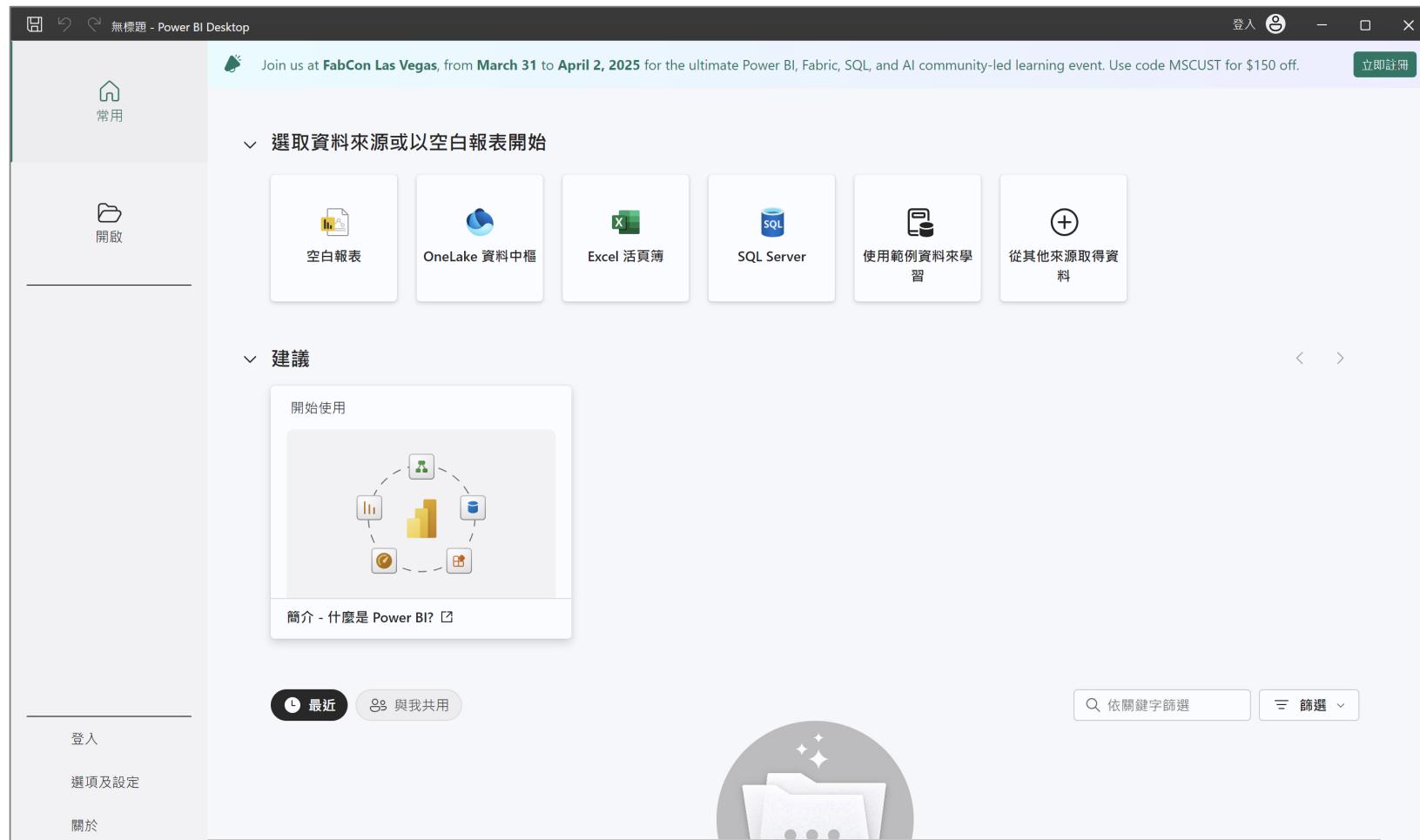
<input type="checkbox"/> 檔案名稱	大小
<input type="checkbox"/> PBIDesktopSetup.exe	504.9 MB
<input checked="" type="checkbox"/> PBIDesktopSetup_x64.exe	544.7 MB

下載 大小總計 : 544.7 MB

安裝 PBIDesktopSetup_x64.exe

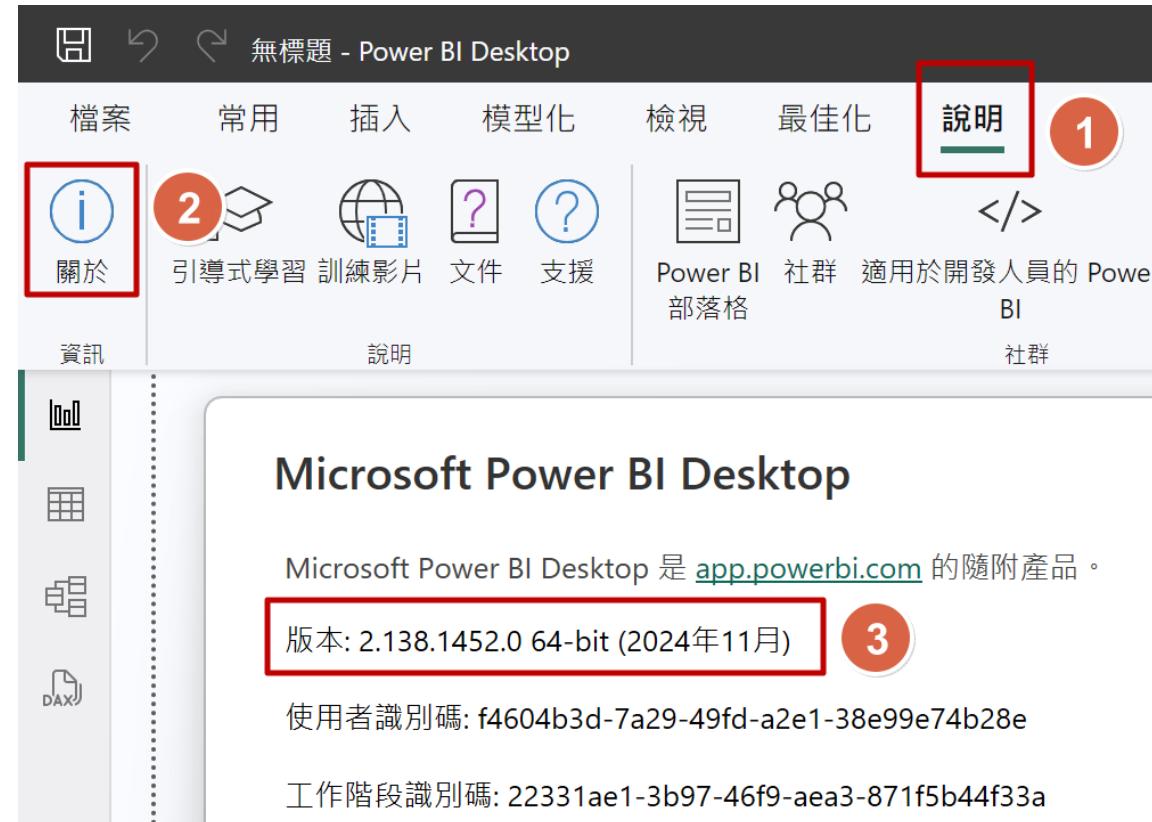


Power BI 啟動視窗



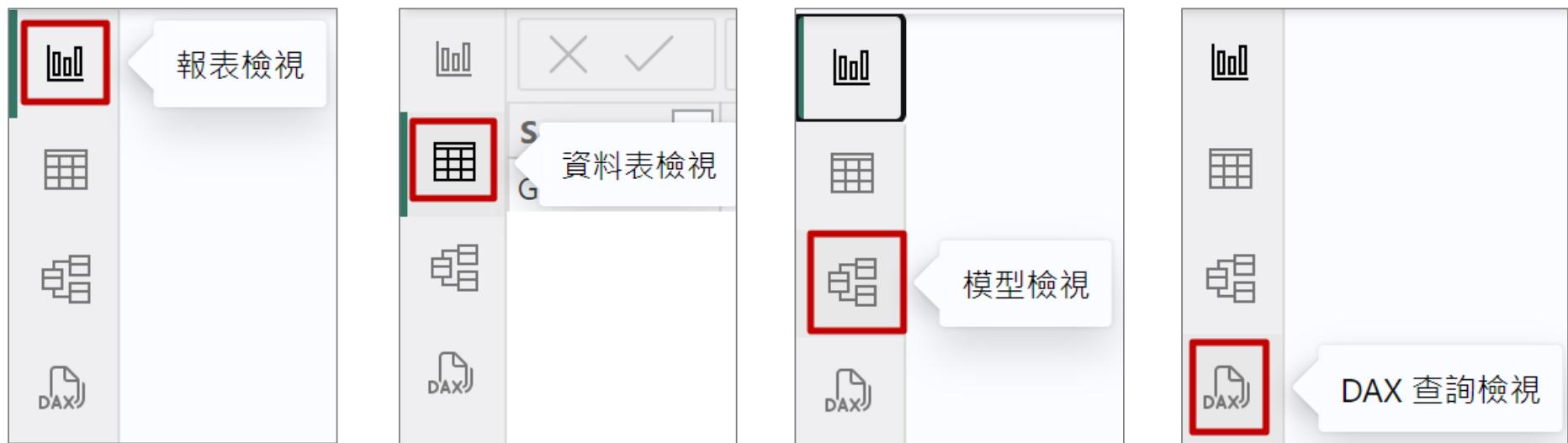
Power BI 版本 2.138

- 說明 \ 關於



Power BI Desktop 四大模組概述

- Report view(Power Viewer) : 報表檢視 - 提供資料視覺化
- Table view (Power Query) : 資料表檢視 - 提供數據工程
- Model view (Power Pivot) : 模型檢視 - 提供關聯式模型
- DAX query view : DAX查詢檢視 – 提供 DAX 功能



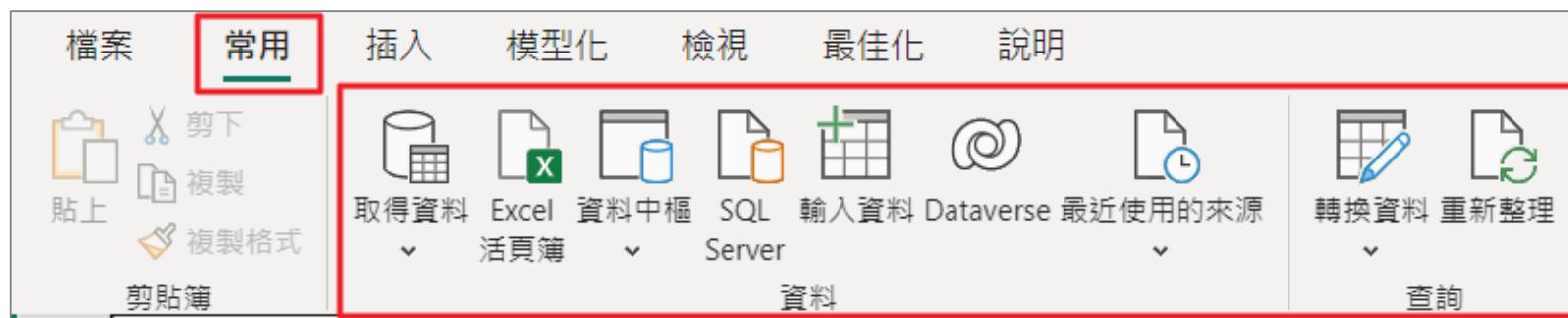
報表檢視

- 插入\頁面、視覺效果、AI視覺效果、Power Platform、元素、走勢圖
- 視覺呈現核心，視覺效果模板提供30種以上圖表類型。
- Power BI市集裡面（Power BI visuals），提供超過200個客製化視覺效果模板檔案。性質除屬於一般商業智慧類型之外，也包含大數據分析（部分須透過R語言繪製設計）的模板檔案。



資料表檢視

- 常用 \ 資料、查詢
- 蒐集資料及清理資料的重要橋樑跟工具。
- 取得異質資料來源的連接後，使用編輯查詢器進行資料清理作業，以利於後續作業。
- 轉換資料 \ Power Query 轉換資料時，可以不用寫程式，過程還會被記錄下來，中途有錯誤時，仍可折返除錯。
- 支援 M語言查詢 (Mashup language query; mashup 混搭，支援多種資料來源)。



模型檢視

- 模型化\關聯性、計算、頁面重新整理、參數、安全性、問與答
- 建立資料模型的核心。利用關聯設定建立資料表之間的關係。
- 同時包括計算主要模組。新增資料行、量值欄位的四則運算及相關量值函數建立等，以利提供設計儀表板上的彈性。



DAX 查詢檢視

- DAX 查詢 (Data Analysis eXpressions) 與 Excel函數相似。
- TotalSales = SUM(Sales[SalesAmount])



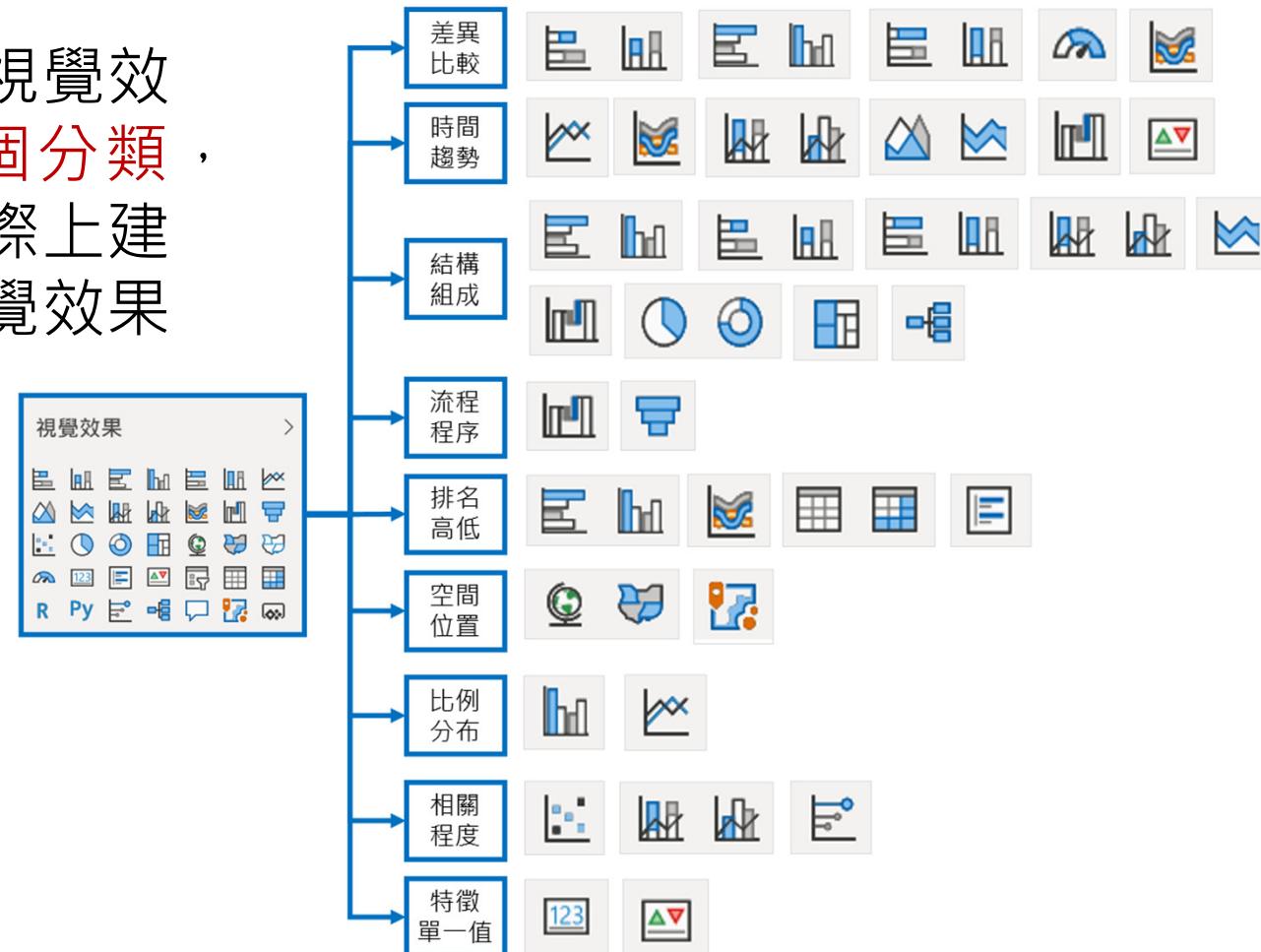
i DAX 查詢將儲存至您的模型 在 Power BI 服務中發佈時，將不會顯示它們。 [深入了解](#) 

 執行  使用變更更新模型 (0)

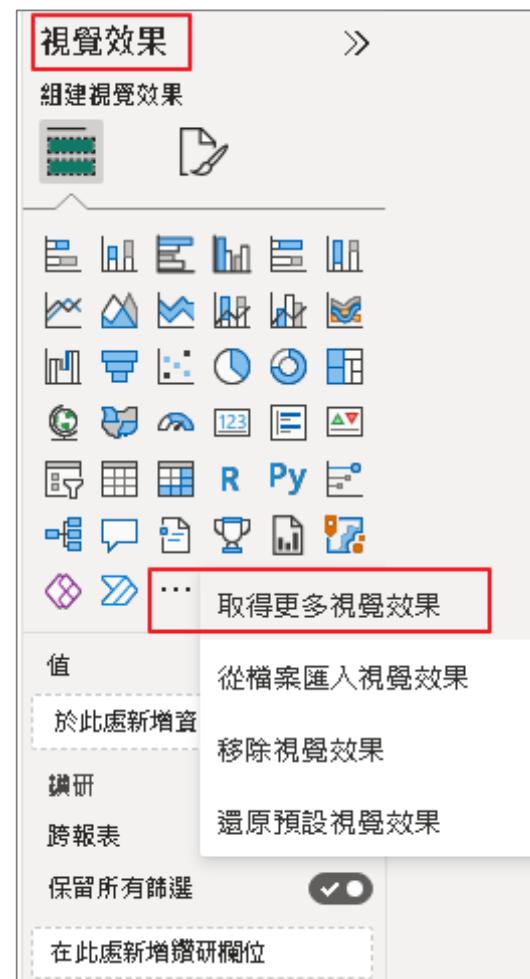
```
1 // 歡迎使用 DAX 查詢檢視! 若要深入了解 DAX 查詢，請至 https://aka.ms/dax-queries。
2 // 以滑鼠右鍵按一下資料窗格中的資料表、資料行或度量，以存取快速查詢，或向 Copilot 尋求撰寫 DAX 的協助。
3 
```

視覺效果模板

- 數據問題可從視覺效果中歸納出**9個分類**，它是一種和實際上建議可以運用視覺效果圖表。



取得更多視覺效果



Power BI 下載視覺效果

Power BI 視覺效果

① 按一下 [新增] 和/或 [下載範例] 並下載視覺效果，即表示您同意視覺效果頁面上提供者的「條款及條件」以及「隱私權原則」，並同意 Microsoft 可以出於交易目的將您的帳戶詳細資料分享給提供者。使用 Microsoft AppSource 必須遵守 [Microsoft 商業市集條款](#) 和 [隱私權聲明](#)。

所有視覺效果 組織視覺效果 AppSource 視覺效果

探索所有可用的視覺效果，以加強您的商務深入解析 [深入了解](#)

篩選依據: 全部 排序依據: 熱門程度

 Gantt  Microsoft Corporation  ★★★★★(158)	 Word Cloud  Microsoft Corporation  ★★★★★(140)	 Advance Card  Bhavesh Jadav  ★★★★★(67)	 Radar Chart  Microsoft Corporation  ★★★★★(129)	 Infographic Designer  Microsoft Corporation  ★★★★★(97)
				

Word Cloud 文字雲

AppSource | 適用於 Power BI visuals 的應用程式

新增

Word Cloud
Microsoft Corporation
4.2 (140)
PBI Certified

概觀 評分與評論

Create a fun visual from frequent text in your data

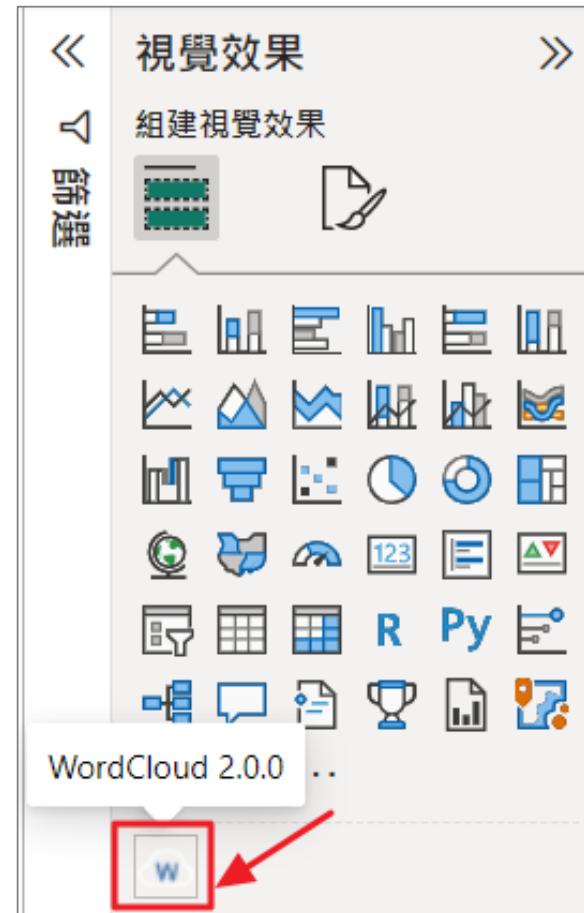
Word Cloud is a visual representation of word frequency and value. Use it to get instant insight into the most important terms in your data. With the interactive experience of Word Cloud in Power BI, you no longer have to tediously dig through large volumes of text to find out which terms are prominent or prevalent. You can simply visualize them as Word Cloud and get the big picture instantly and user Power BI's interactivity to slice and dice further to uncover the themes behind the text content. This visual also puts you in control on the appearance of the word cloud, be it the size or usage of space and how to treat the data. You can choose to break the words in the text to look for the frequency word or keep word break off to project a measure as a value of the text. You can also enable stop words to remove the common terms from the word cloud to avoid the clutter. By enabling rotation and playing with the angles allowed, you can become very creative with this visual. Optionally you can also use a measure to provide weightage to the text. If none provided, it will simply use the frequency. Check out the formatting pane for more options. This is an open source visual. Get the code from GitHub: <https://github.com/Microsoft/PowerBI-visuals-wordcloud>

視覺效果功能

在視覺效果中 Power BI 說明



Word Cloud 新增完成



M語言查詢 vs. DAX查詢

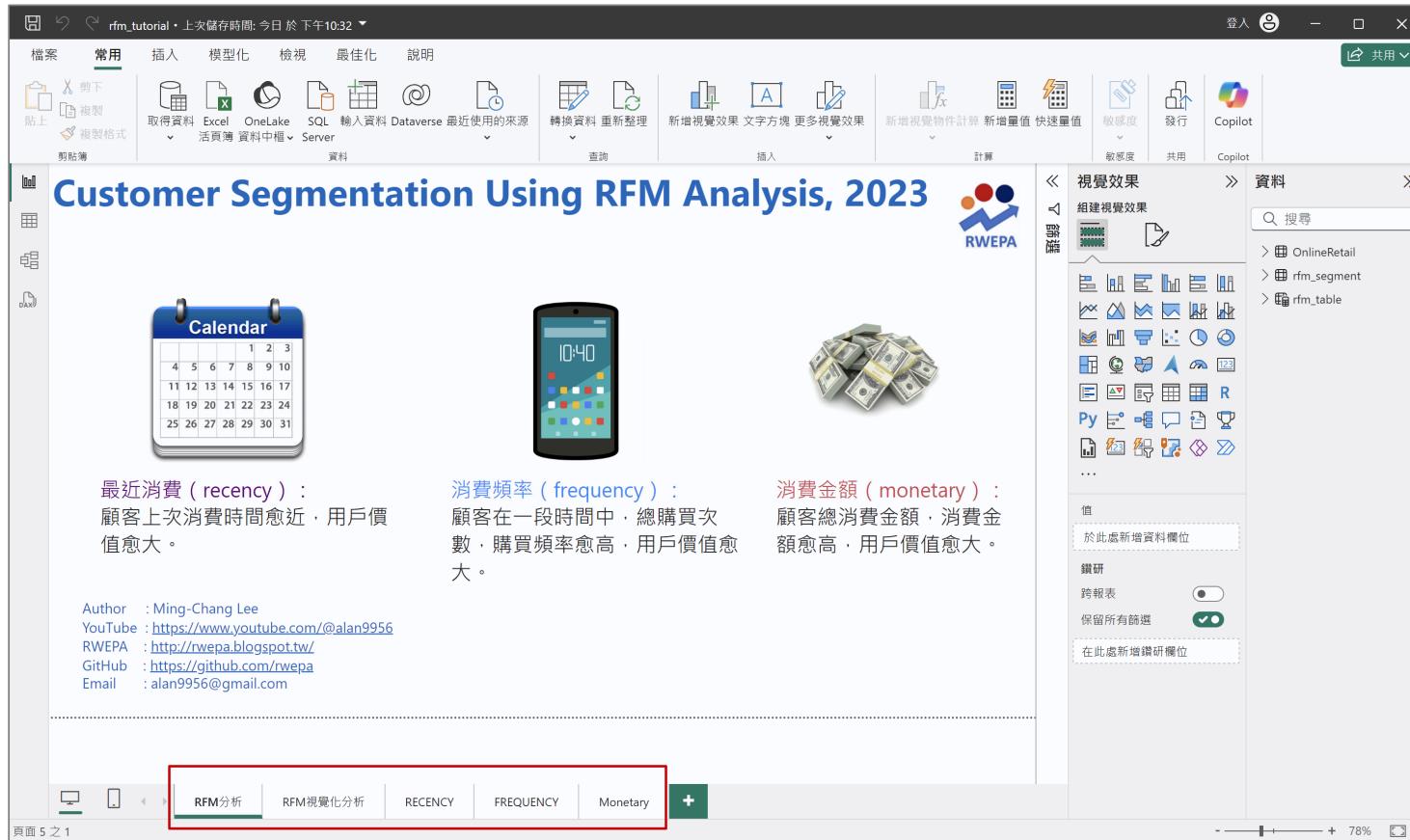
- M 語言查詢更類似於 C# 中的 SQL 或 LINQ (語言集成查詢) 命令。
- M 語言用於在查詢編輯器中管理和過濾數據。
- DAX 查詢主要在數據視圖中執行。
- DAX 查詢更類似於 Excel。
- M 語言用於在查詢編輯器中預處理數據，而 DAX 用於在數據加載到數據視圖模型後分析數據。
- M 語言用於數據工程（有時稱為數據準備），DAX 用於數據分析。
- M (i.e. in data engineering) and in DAX (i.e. in data analysis)
 - <https://learn.microsoft.com/zh-tw/dax/dax-function-reference>

M語言查詢 vs. DAX查詢(續)

- M語言適用於查詢階段，相當於做了一個特定的模具，將符合條件的數據導入到Power BI當中，之後這些數據將會用來建立各種報表。
- 通過M語言過濾掉無用數據或將特定數據合併整理，可以減輕後續Power BI在分析數據上對CPU和RAM的壓力，提高效率。
- DAX語言是對導入後的數據進行分析，通過DAX函數計算得到的結果可以被直接用來生成報表（視覺化），並且通過添加過濾條件，DAX語言可以便捷的獲取某個特定區間內的結果。更適用於數據分析的彙總，加成，區間計算，特定條件求值等需求。

Power BI - RFM視覺化分析範例

- https://github.com/rwepa/DataDemo/blob/master/rfm_tutorial.pbix



The screenshot shows the Power BI desktop interface with a report titled "Customer Segmentation Using RFM Analysis, 2023". The report contains three main visualizations: a calendar icon with the text "最近消費 (recency) : 顧客上次消費時間愈近，用戶價值愈大。", a smartphone icon with the text "消費頻率 (frequency) : 顧客在一段時間中，總購買次數，購買頻率愈高，用戶價值愈大。", and a stack of money with the text "消費金額 (monetary) : 顧客總消費金額，消費金額愈高，用戶價值愈大。". The ribbon at the top has tabs for "RFM分析", "RFM視覺化分析", "RECENTY", "FREQUENCY", and "Monetary". The "RFM視覺化分析" tab is selected. The Power BI logo and the text "RWEPA" are visible in the bottom right corner.

3. Tableau簡介

Tableau 視覺應用介紹

- Tableau Software 美國互動式資料視覺化軟體公司，由史丹佛大學電腦科學系 [Christian Chabot](#) (VizQL技術), [Chris Stolte](#)(資料庫), [Pat Hanrahan](#) (圖形引擎) 於2003年1月在加利福尼亞州山景城創立。該公司目前總部位於美國華盛頓州西雅圖市，專注於商業智慧。
- 2019年8月1日，Salesforce.com 以157億美元收購了Tableau。



Christian Chabot



Chris Stolte



Pat Hanrahan

Tableau – 視覺化創新應用

- 圖形化操作介面, 使用滑鼠拖曳與選取
- 多樣化的資料(資料庫)連結
- 輕鬆建立視覺化圖表
- 資料過濾與資料層級剖析
- 快速分享研究結果
- API應用 (**R, Python, JavaScript, C, C++, Java, ...**)
 - 使用R統計分析功能
 - 使用R建立機器學習模型
 - 使用Python 建立深度學習模型
 - 將R/Python模型回傳至 Tableau , 進行進階視覺化應用(Power BI 亦可完成)
 - 支援 Windows and macOS 作業系統 (**Power BI 僅支援 Windows**)

Tableau 產品

- Tableau Desktop (14天適用版)
- Tableau Public (免費的平台)
- Tableau Server
- Tableau Cloud
- Tableau Prep (清理與組織資料)

Tableau 下載-方法1

- 產品版本 <https://www.tableau.com/zh-tw/products/desktop>



The screenshot shows the Tableau Desktop product page on the salesforce website. The top navigation bar includes links for '為什麼選擇 Tableau' (Why Choose Tableau), '產品' (Products), '解決方案' (Solutions), '資源' (Resources), and '合作夥伴' (Partners). On the right, there are buttons for '立即購買' (Buy Now) and '立即試用' (Try Now). The main content area features a heading 'Tableau Desktop' and the tagline '以業務步調取得見解' (Get insights at your own pace). A large orange button labeled '開始免費試用' (Start Free Trial) is highlighted with a red box and an arrow. To the right, there is a visual representation of the Tableau interface showing a data visualization with a 'Marks' shelf, a 'Clusters' section, and a 'Show Me' feature.

開始免費試用
【下載最新版本】

Tableau 下載-方法2 【★★★推薦方法】

- 選取下載 Tableau Desktop 網頁

<https://www.tableau.com/zh-tw/support/releases>



The screenshot shows the 'PRODUCT RELEASES' section of the Tableau website. On the left, there's a sidebar with links: Desktop (highlighted with a red box and circled '1'), Prep, Server, Cloud Bridge, Advanced Management, Mobile Blackberry, Intune, and Workspace ONE. The main area lists releases by year:

年份	版本	發佈日期
2024.3	2024.2.5	已發佈 21/11/2024
2024.2	2024.2.4	已發佈 24/10/2024
	2024.2.3	已發佈 24/09/2024
	2024.2.2	已發佈 20/08/2024
	2024.2.1	已發佈 24/07/2024

A yellow callout box highlights the following points:

- 考量14天試用版的使用
- 安裝新版本後, 可能無法安裝舊版本
- 先選取【2024.2.1版本】, 到期後再選取【2024.2.2版本】

Tableau 安裝



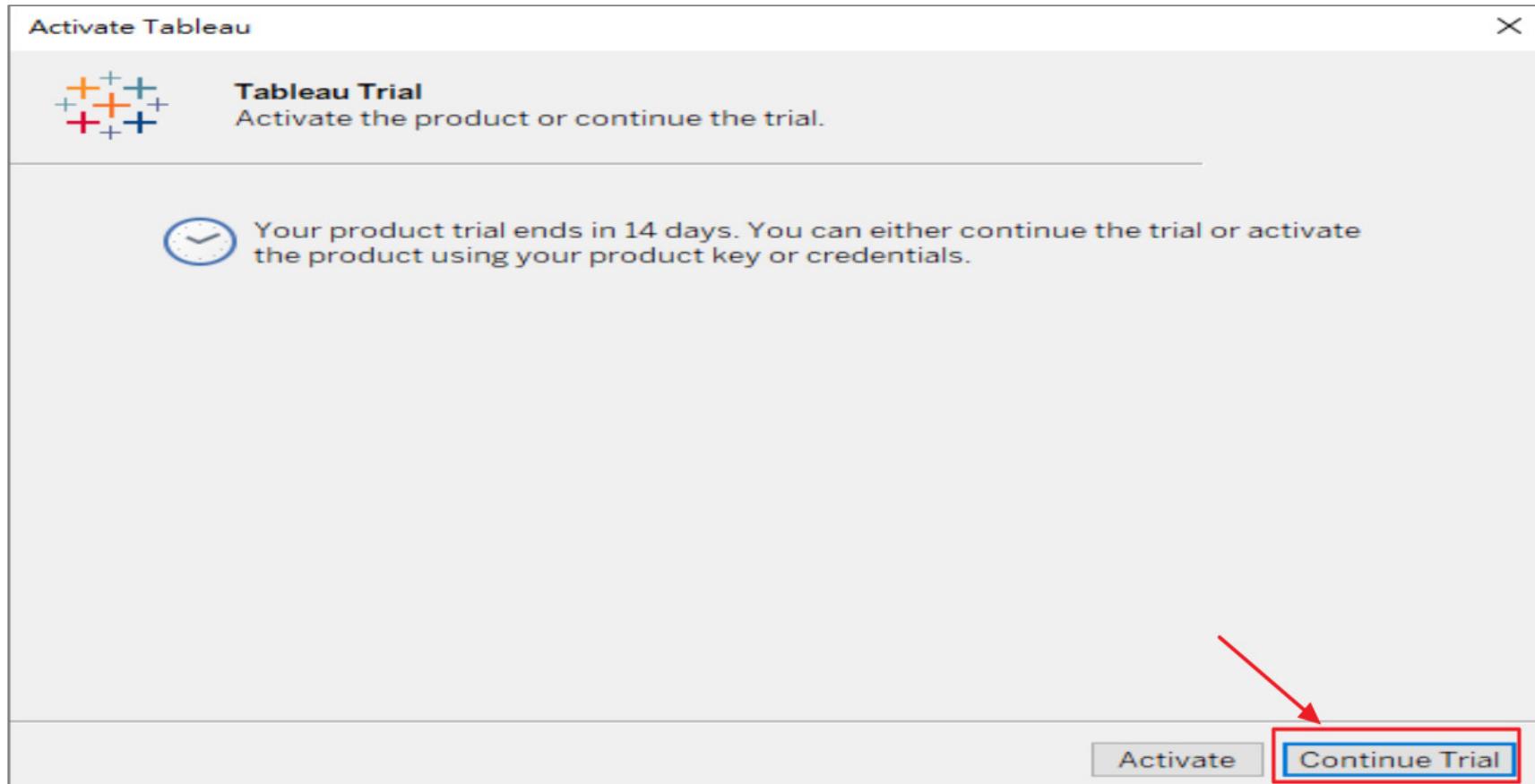
允許此 App 變更您的裝置



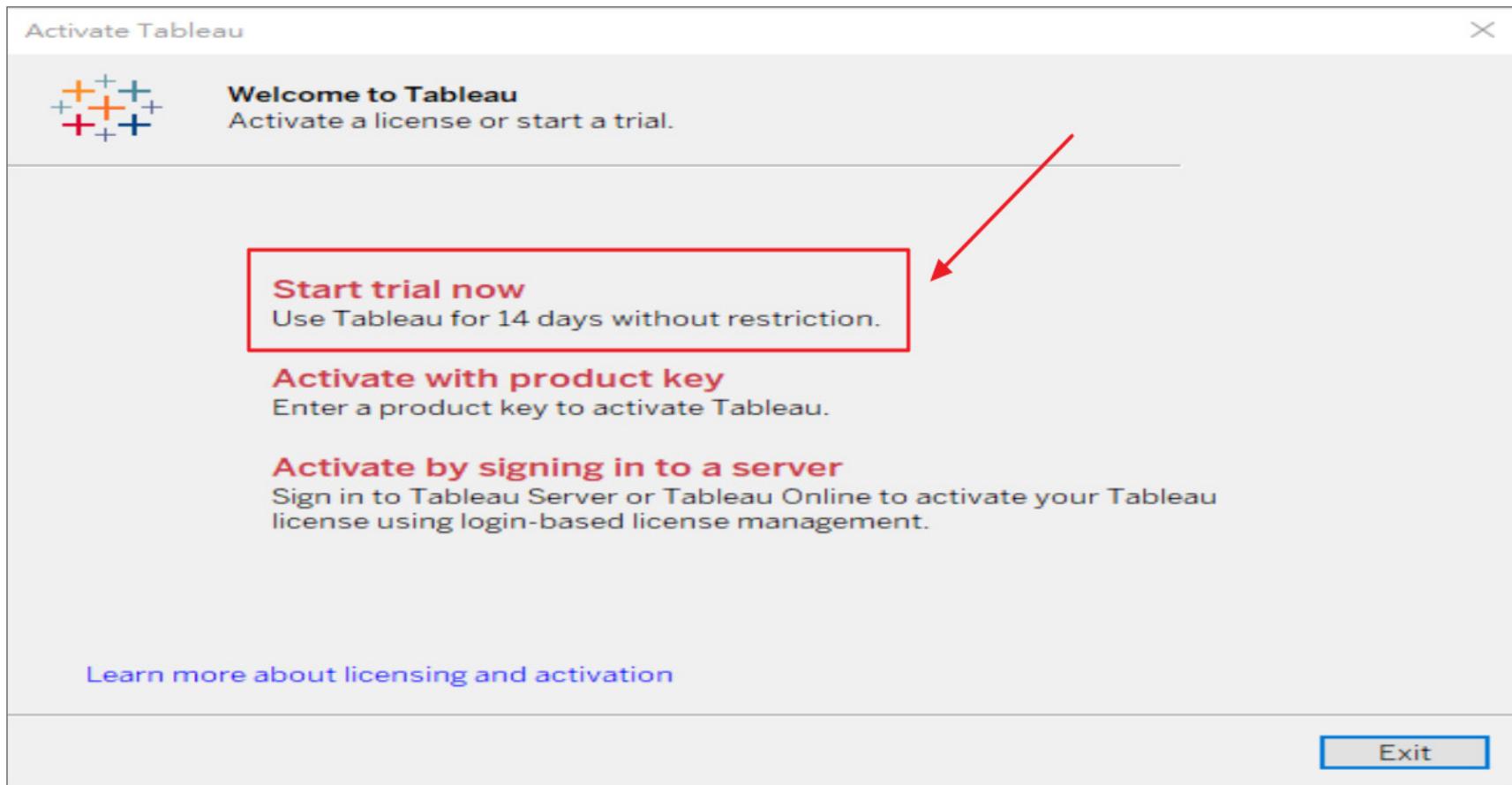
安裝畫面



安裝完成



Start trial now



Registration 產品註冊

Activate Tableau

 **Registration**
Complete all fields for the registered user.

First Name <input type="text" value="ALAN"/>	Last Name <input type="text" value="LEE"/>	Organization <input type="text" value="WWW"/>
Email <input type="text" value="[REDACTED]@[REDACTED].com"/>	Job Title <input type="text" value="user"/>	Department <input type="text" value="Engineering/Develop..."/>
Phone <input type="text" value="0123456789"/>	State/Province <input type="text" value="Taiwan"/>	Company Size <input type="text" value="1 - 20 employees"/>
<p>Your personal data may be used by Salesforce for marketing purposes. You may unsubscribe from Salesforce marketing communications at any time as described in the Privacy Statement.</p> <p>By submitting this form, you confirm that you agree to the storing and processing of your personal data by Salesforce as described in the Privacy Statement.</p>		
 <input type="button" value="Register"/>		

產品註冊完成

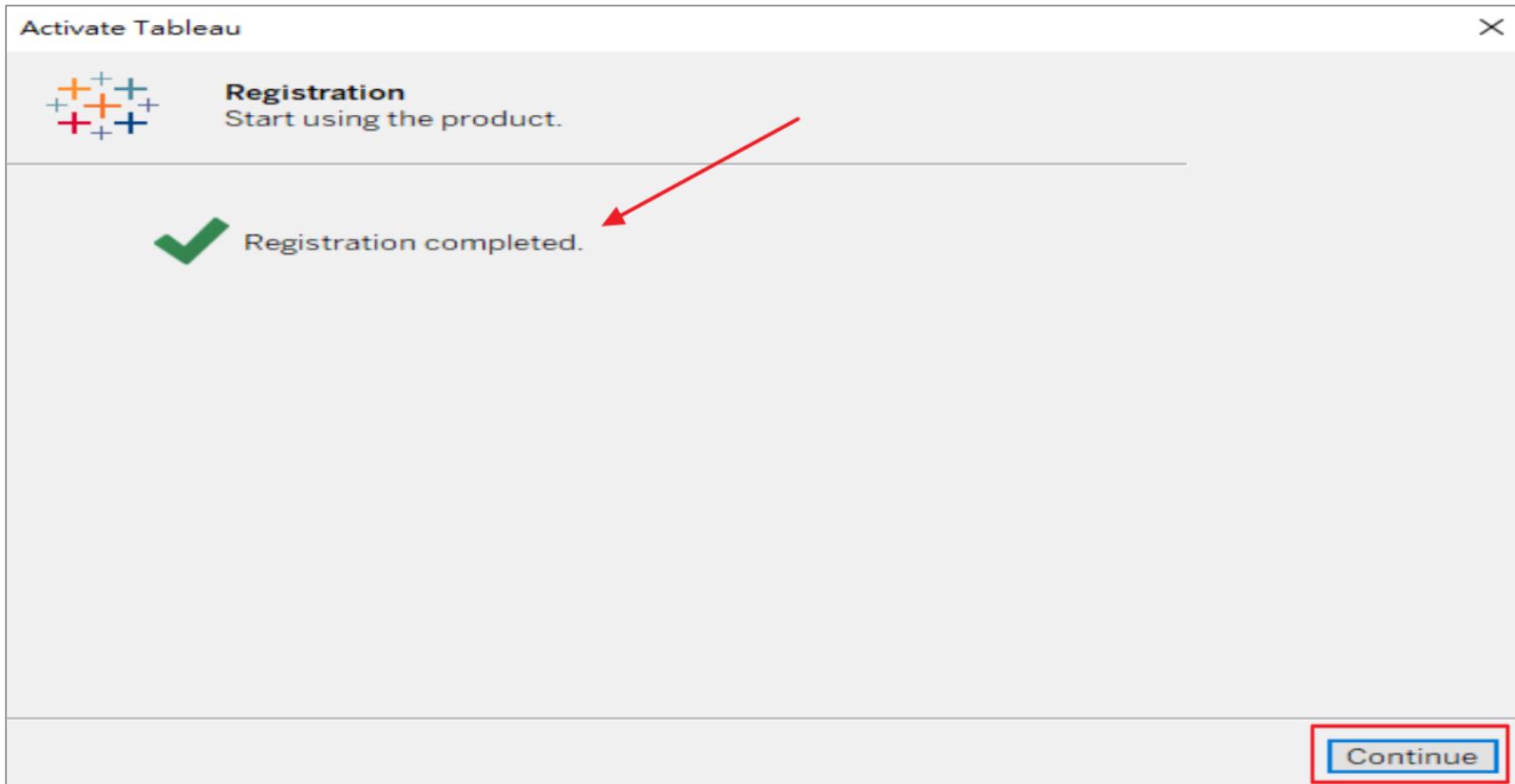


Tableau 預設畫面-英文版

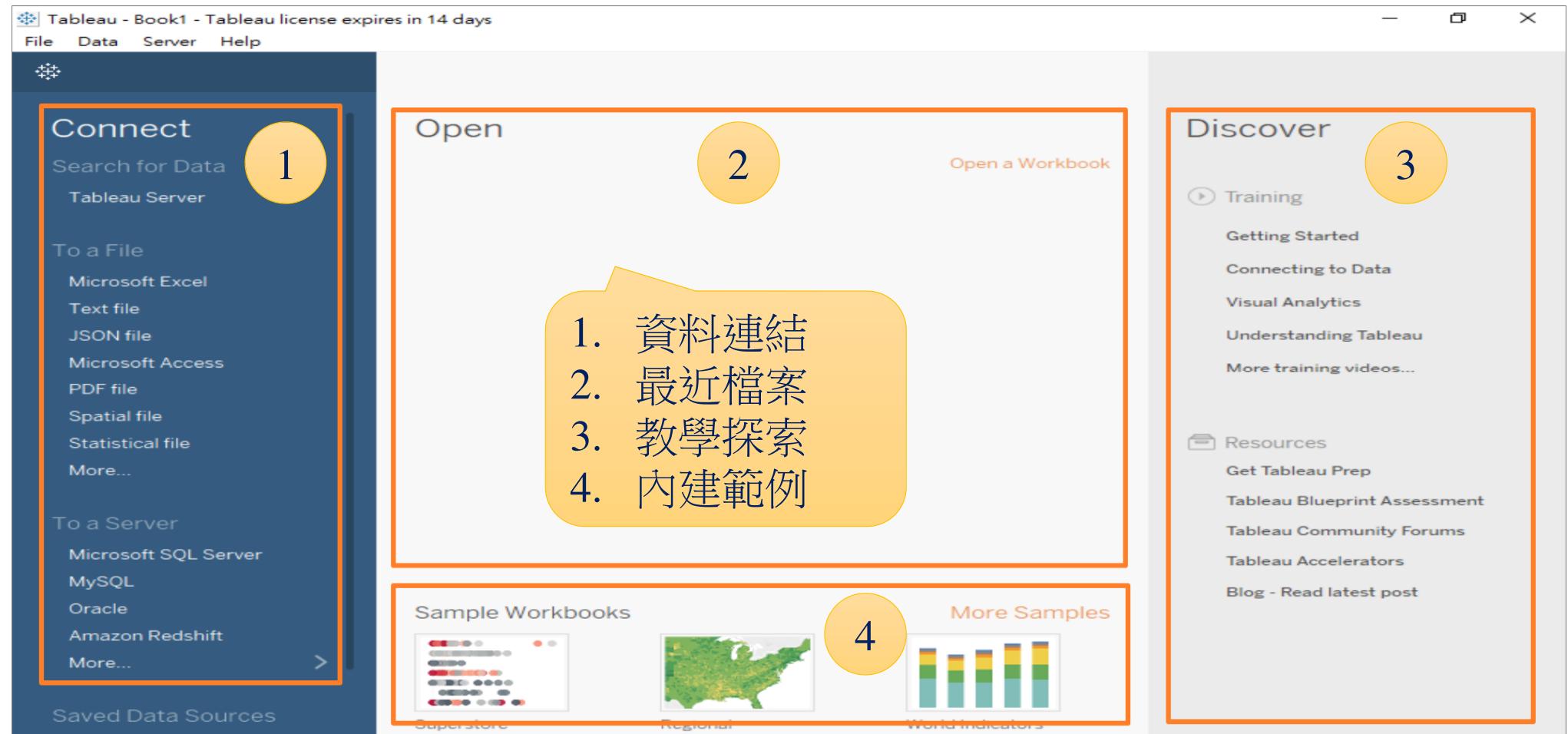


Tableau 預設畫面-中文版

The screenshot shows the Tableau 2022.2 welcome screen. A yellow callout box highlights the four main sections:

- 資料連結 (Data Sources) - Located in the top-left corner of the main workspace.
- 最近檔案 (Recent Files) - Located in the center of the main workspace.
- 教學探索 (Learning Exploration) - Located in the top-right corner of the main workspace.
- 內建範例 (Built-in Examples) - Located at the bottom center of the main workspace.

Each section is highlighted with a red box and a yellow circle containing a number 1, 2, 3, or 4. The main workspace also features a red border around the central area.

1. 資料連結

2. 最近檔案

3. 教學探索

4. 內建範例

開啟 (Open) - Located at the top left of the main workspace.

開啟工作簿 (Open Workbook) - A small window titled "歡迎使用" (Welcome) with the message: "您的 14 天試用版已啟動。可透過以下這些很好的方法瞭解 Tableau。"

探索 (Explore) - Located in the top-right corner of the main workspace.

資源 (Resources) - Located in the bottom-right corner of the main workspace.

Tableau 2022.2 現已推出 (Tableau 2022.2 is now available) - Located in the bottom-right corner of the main workspace.

更多範例 (More Examples) - Located in the bottom-right corner of the main workspace.

立即升級至 2022.2 (Upgrade to 2022.2 now) - Located in the bottom-right corner of the main workspace.

連線 (Connections) - Located in the top-left corner of the main workspace.

開啟 (Open) - Located in the top-left corner of the main workspace.

探索 (Explore) - Located in the top-right corner of the main workspace.

資源 (Resources) - Located in the bottom-right corner of the main workspace.

Tableau 2022.2 現已推出 (Tableau 2022.2 is now available) - Located in the bottom-right corner of the main workspace.

立即升級至 2022.2 (Upgrade to 2022.2 now) - Located in the bottom-right corner of the main workspace.

連線 (Connections)

- 搜尋資料
- Tableau Server
- 到檔案

 - Microsoft Excel
 - 文字檔
 - JSON 檔案
 - Microsoft Access
 - PDF 檔案
 - 空間檔案
 - 統計檔案
 - 更多...

- 到伺服器

 - Microsoft SQL Server
 - MySQL
 - Oracle
 - Amazon Redshift
 - 更多...

- 已儲存資料來源

 - Sample - Superstore
 - 世界指標
 - 範例 - 超級市場

開啟 (Open)

開啟工作簿 (Open Workbook)

歡迎使用 (Welcome)

您的 14 天試用版已啟動。可透過以下這些很好的方法瞭解 Tableau。

探索 (Explore)

- 訓練
- 快速入門
- 連線至資料
- 視覺分析
- 認識 Tableau
- 更多訓練影片...

資源 (Resources)

- 取得 Tableau Prep
- Tableau Blueprint 評估
- Tableau 社群論壇
- Tableau 加速器
- 部落格 - 閱讀最新文章

Tableau 2022.2 現已推出 (Tableau 2022.2 is now available)

針對資料&部署取得更深入的洞察資料

立即探索 →

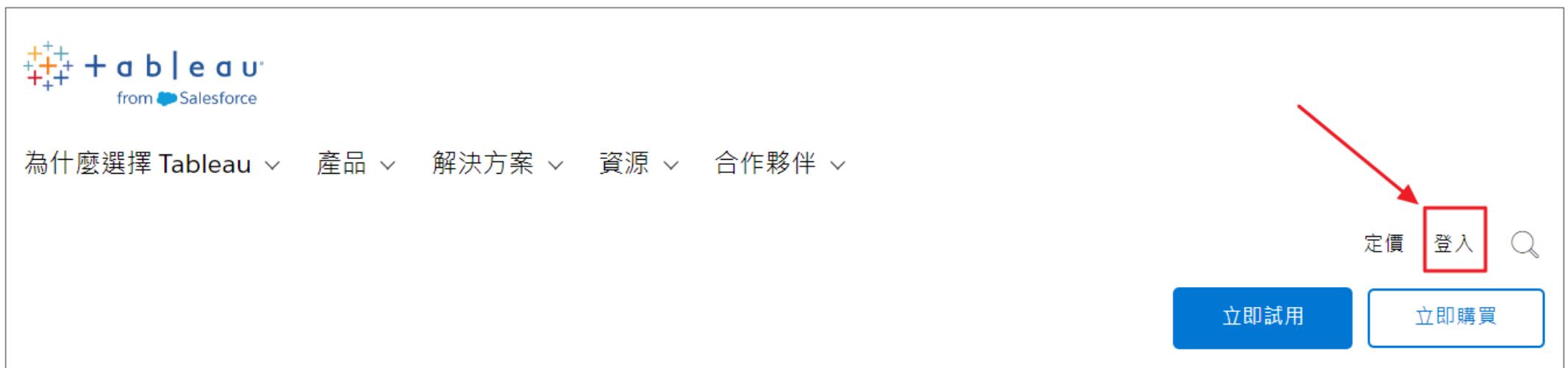
更多範例 (More Examples)

立即升級至 2022.2 (Upgrade to 2022.2 now)

立即升級至 2022.2 (Upgrade to 2022.2 now)

Tableau Public 線上版(免安裝版)

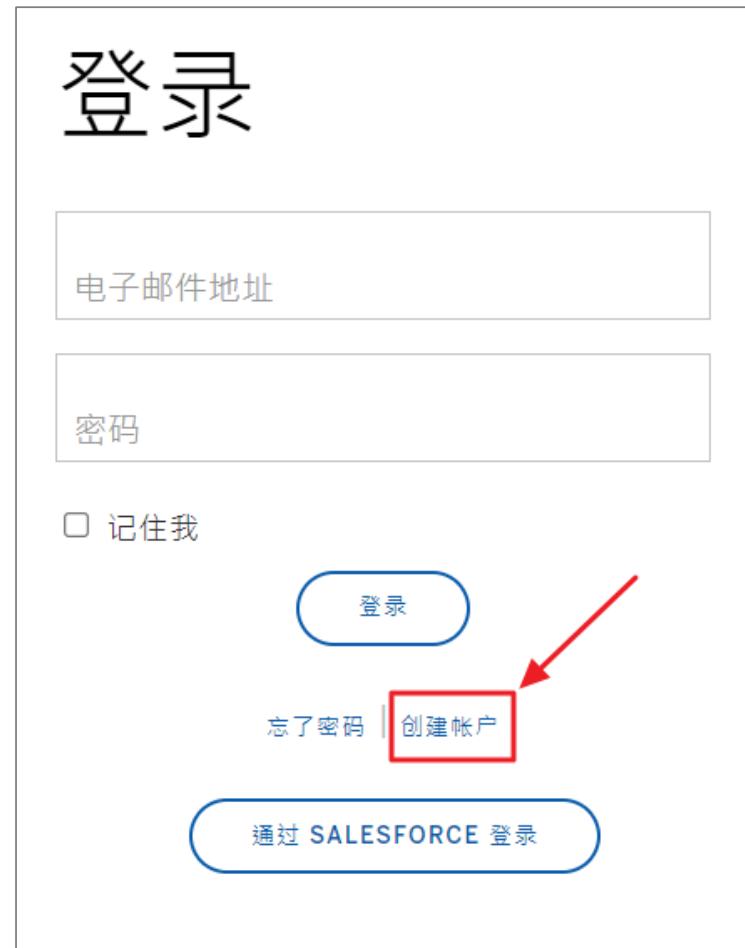
- <https://www.tableau.com/zh-tw/community/public>
- 選取 [登入]



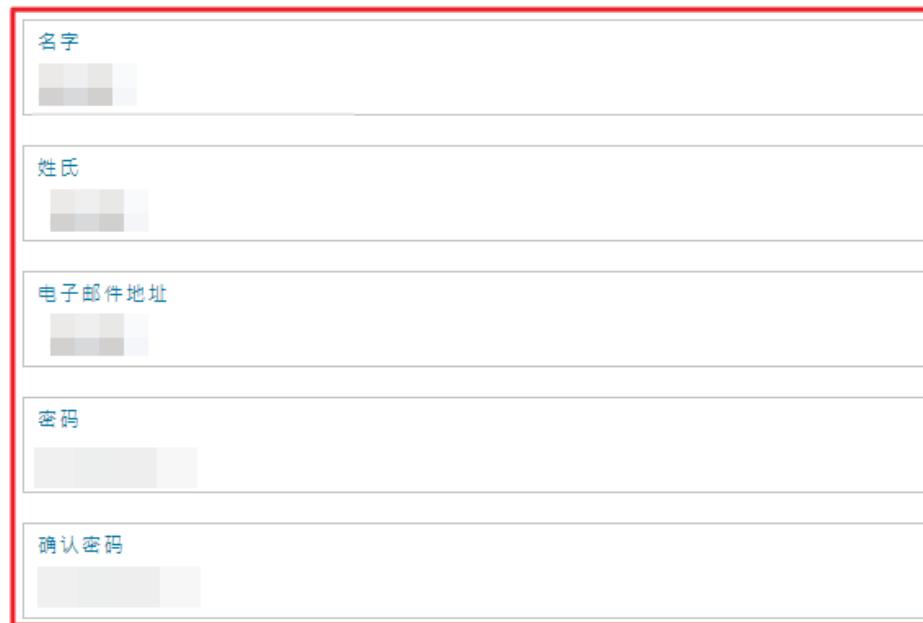
登入 Tableau Public



登錄 \ 創建帳戶



創建我的帳戶



名字
姓氏
电子邮件地址
密码
确认密码

密码区分大小写，长度为 8-100 个字符。要求至少包含一个数字 (0-9)、一个大写字母、一个小写字母，以及下列字符之一: ! " \ # \$ % & ' () * + , - . / : ; < = > ? @ [] ^ _ { | } ~ 。

创建我的帳戶

註冊完成通知，登入電子郵件



感谢您注册。您将收到一封电子邮件，其中包含用于确认此注册的链接。您必须按照电子邮件中的说明操作，才能激活您的 Tableau 帐户。

快速链接

- [Tableau 主页](#)
- [Tableau 社区](#)
- [Blueprint 评估](#)

電子郵件 - Tableau Software 帳戶激活

Tableau Software 帳戶激活 ► 收件匣 ×

Tableau <reghelp@tableau.com>
寄給我 ▾

 + a b | e a u®

感谢您注册 Tableau。我们很高兴您成为了 Tableau 社区的一员，并希望您经常回来，从我们的论坛、培训视频、用户群组和其他资源中受益。

若要激活您的帳戶，请单击下方 URL :

为安全起见，上方链接将在 48 小时后过期。如有必要，您可以在此处申请新链接：
<https://id.tableau.com/resendActivation>

如果您未提出此请求，请忽略此电子邮件。

为确保您能够收到未来的通信，请您将“tableau.com”添加到您的安全发件人列表。

再次感谢您的注册。

Tableau 团队

© 2003-2023 Tableau Software, LLC, a Salesforce Company. 保留所有权利!

點選 URL



歡迎使用 Tableau Public \ 登入



The screenshot shows the Tableau Public homepage. At the top left is the Salesforce logo. Below it are the Tableau Public logo, a "建立" (Create) dropdown menu, and a "了解" (Learn) link. A red arrow points to the "登入" (Log In) button on the right. A search icon is also present. A banner at the top asks if you want to boost your skills by connecting with the Tableau community. The main content features a large, colorful network visualization titled "China - by far the biggest consumer in the world." The visualization shows connections between various countries, with China at the center. Below the visualization, text reads "The pace at which Qatar is...". On the left, there's a large "歡迎使用 Tableau Public" heading and a subtext explaining it's a free platform for exploring and sharing data visualizations. At the bottom left is a blue "註冊Tableau Public" button, and at the bottom right is a "深入瞭解" link.

登錄 Tableau Public 帳號



Tableau Public 需要更多信息

Tableau Public 需要更多信息

选择您所在的国家或地区。

Land/Region

Taiwan

我同意条款、数据策略和社区代码。

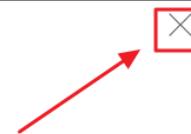
提交此表格即表示您确认并同意您的个人信息可能被传输到中华人民共和国境外的服务器，存储在该服务器上，并在该服务器上进行处理。Salesforce 将会按照[隐私声明](#)处理您的个人信息。

完成登录

* 这些偏好设置将被储存在您的账户中，后续登录时不会再提示。

您為何使用 Tableau Public? \ 按 X 關閉

您為何使用 Tableau Public ?



透過分享您計劃如何在我們的平台上度過您的時間，幫助我們改進 Tableau Public，同時專注於對您至
關重要的改進。

請選取所有適用的選項。

學習資料技能

處理視覺效果、完成課程作業或提高
Tableau 技能。

職業發展

尋找工作或合格的候選人。

尋找靈感

瀏覽其他人建立的內容或查找用於探索
的資料。

加入社群

參與社群專案、社群論壇或與
#DataFam 建立聯繫。

展示我的作品

展示您的作品集，或與他人分享您的作
品。

只是好奇

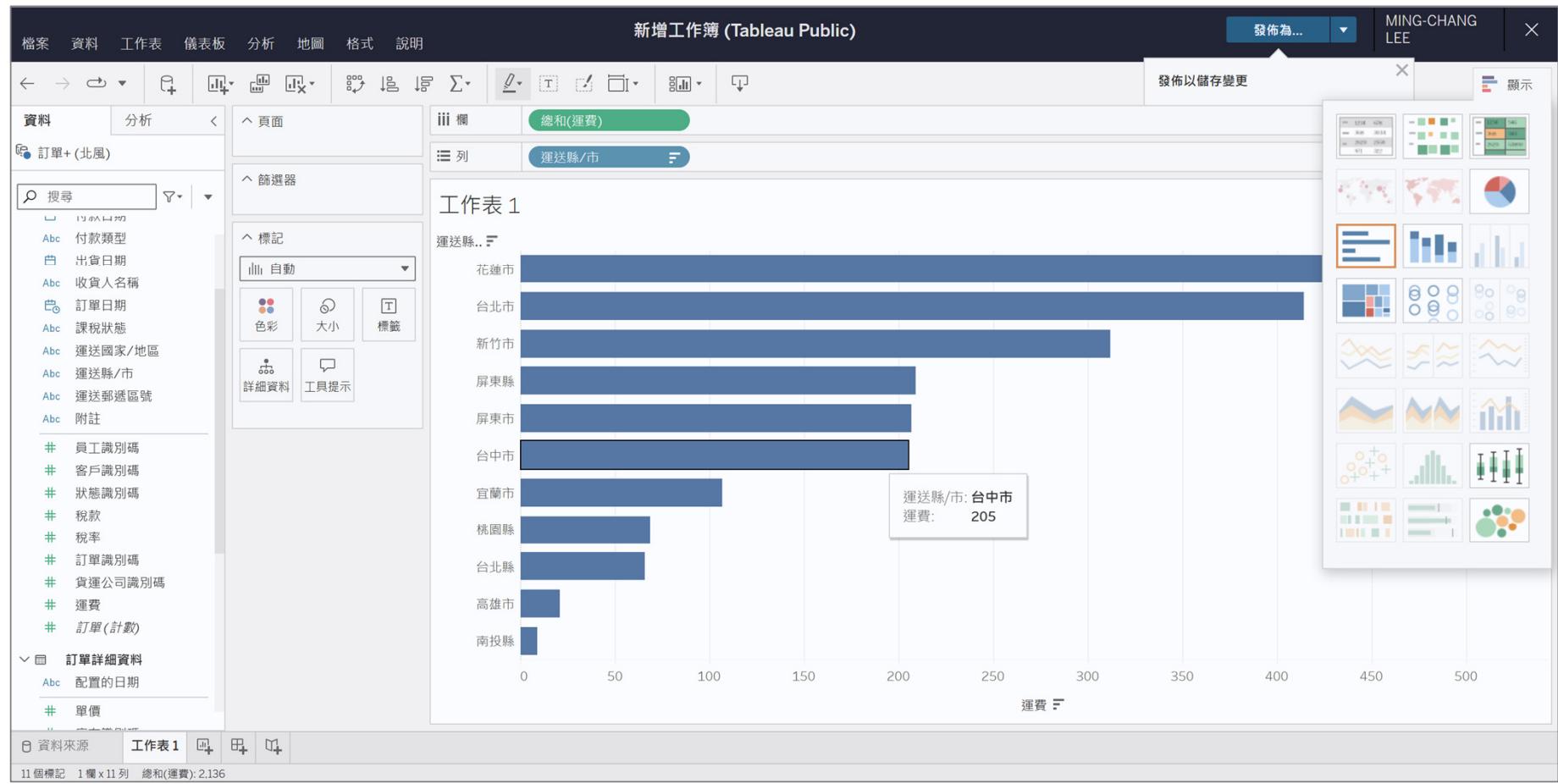
我只是來這裡四處看看。

Tableau Public 啟動畫面

- 二大功能：
 - 建立 \ Web 製作
 - 下載 Tableau Desktop 公共版



Tableau Public – 完成圖



4. R簡介

參考 第2章 R語言簡介與基本運算功能

5. Python簡介

Anaconda 特性

- Anaconda是一個免費、易於安裝/管理並支援Python語言
- 支援 8000個以上的資料科學與機器學習模組 (package)
- 支援模組的下載, 安裝, 更新
- 支援 Spyder (支援 Python IDE)
- 支援 jupyter-notebook
- 支援 Windows、Mac OS和Linux

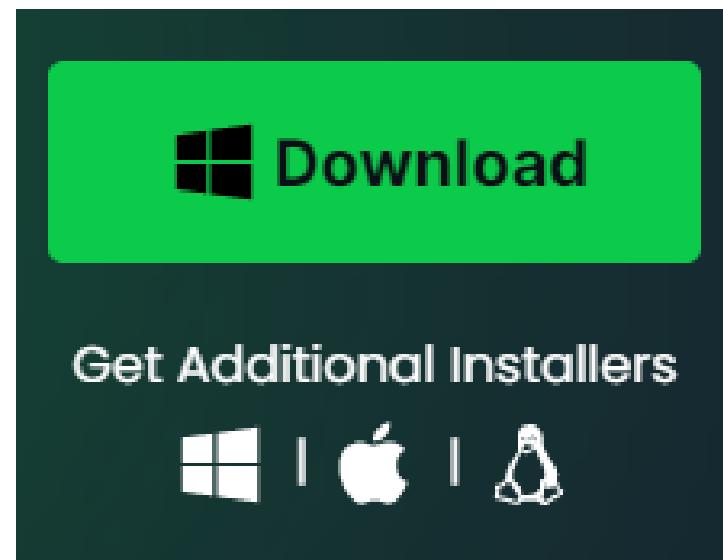
Anaconda 特性 (續)



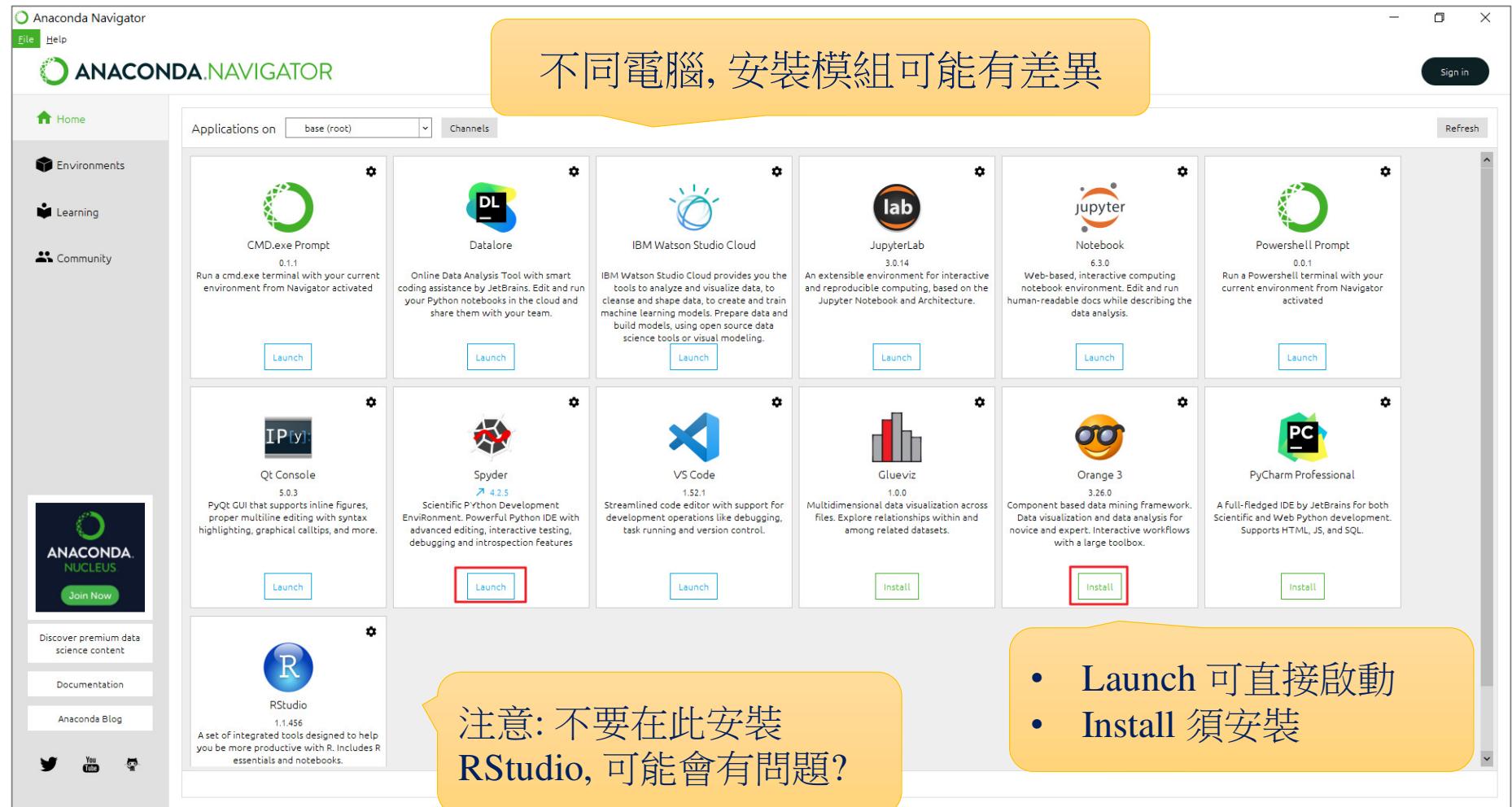
source: <https://www.anaconda.com/distribution/>

Anaconda 下載

- <https://www.anaconda.com/>



Anaconda Navigator



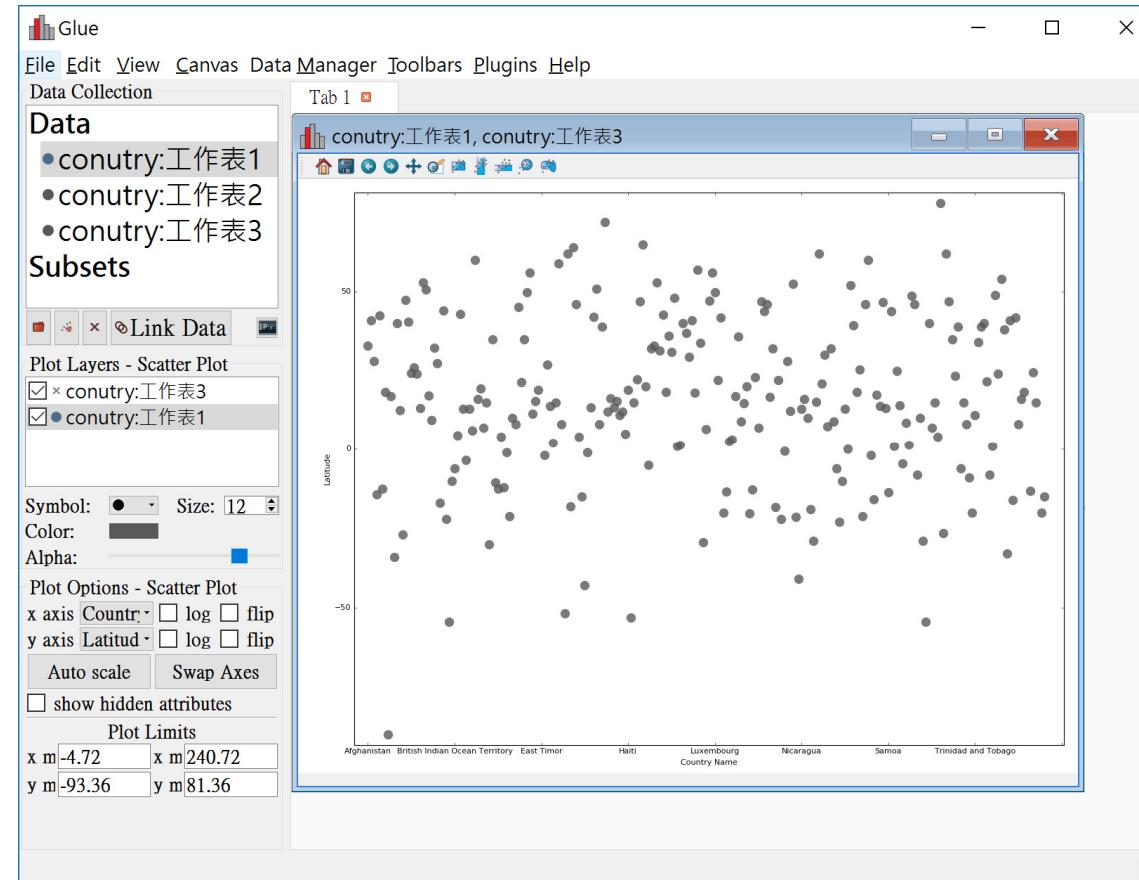
不同電腦, 安裝模組可能有差異

注意: 不要在此安裝 RStudio, 可能會有問題?

- Launch 可直接啟動
- Install 須安裝

Application	Description	Action
CMD.exe Prompt	Run a cmd.exe terminal with your current environment from Navigator activated.	Launch
Datalore	Online Data Analysis Tool with smart coding assistance by JetBrains. Edit and run your Python notebooks in the cloud and share them with your team.	Launch
IBM Watson Studio Cloud	IBM Watson Studio Cloud provides you the tools to analyze and visualize data, to cleanse and shape data, to create and train machine learning models. Prepare data and build models, using open source data science tools or visual modeling.	Launch
JupyterLab	An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.	Launch
Jupyter Notebook	Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.	Launch
Powershell Prompt	Run a Powershell terminal with your current environment from Navigator activated	Launch
IPyConsole	PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.	Launch
Spyder	Scientific Python Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features.	Launch
VS Code	Streamlined code editor with support for development operations like debugging, task running and version control.	Launch
Glueviz	Multidimensional data visualization across files. Explore relationships within and among related datasets.	Install
Orange 3	Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.	Install
PyCharm Professional	A full-fledged IDE by JetBrains for both Scientific and Web Python development. Supports HTML, JS, and SQL.	Install
RStudio	A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.	Launch

glueviz 視覺化



jupyter-notebook



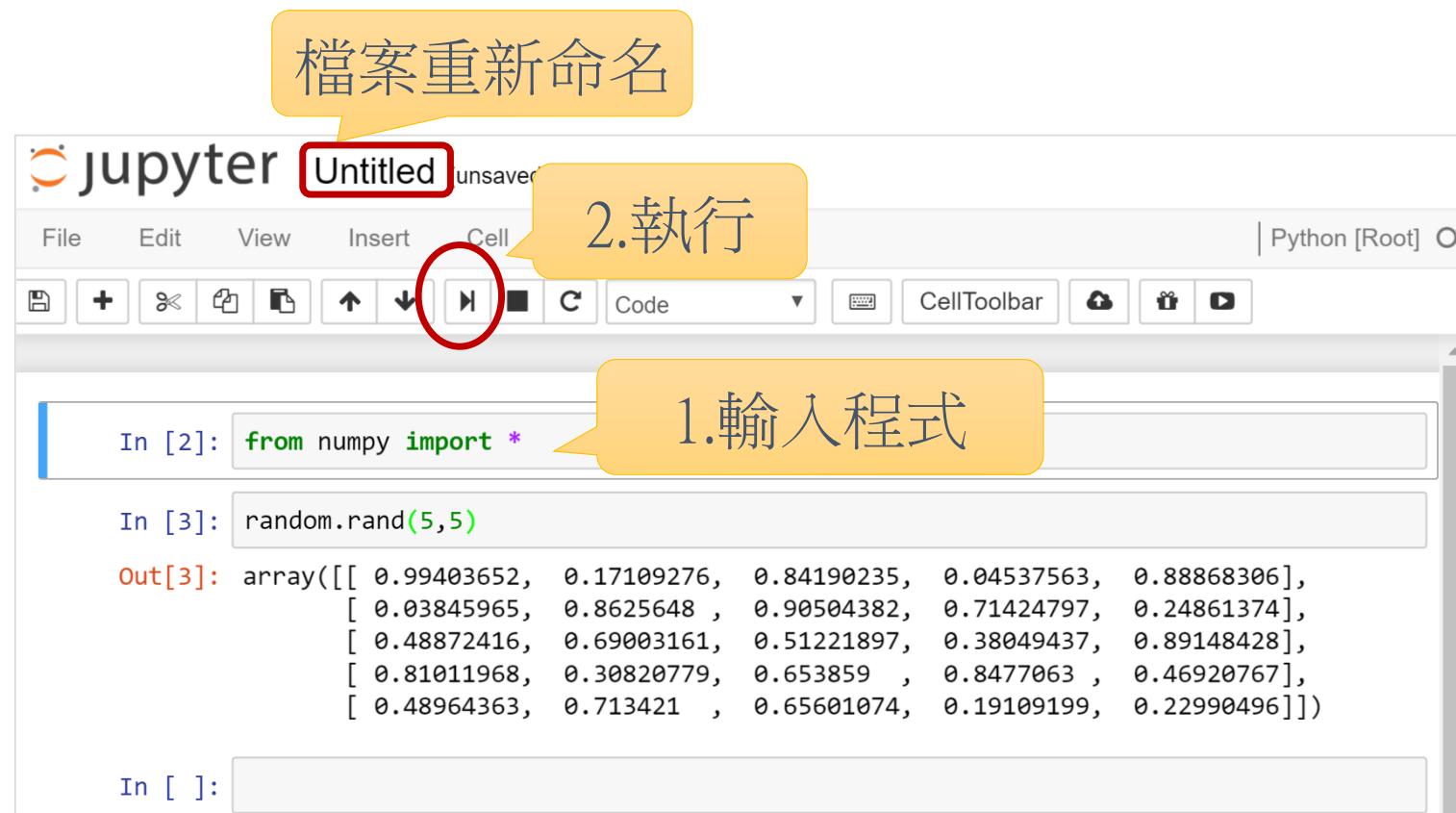
實作
練習

jupyter notebook (續)

檔案重新命名

2. 執行

1. 輸入程式



In [2]: `from numpy import *`

In [3]: `random.rand(5,5)`

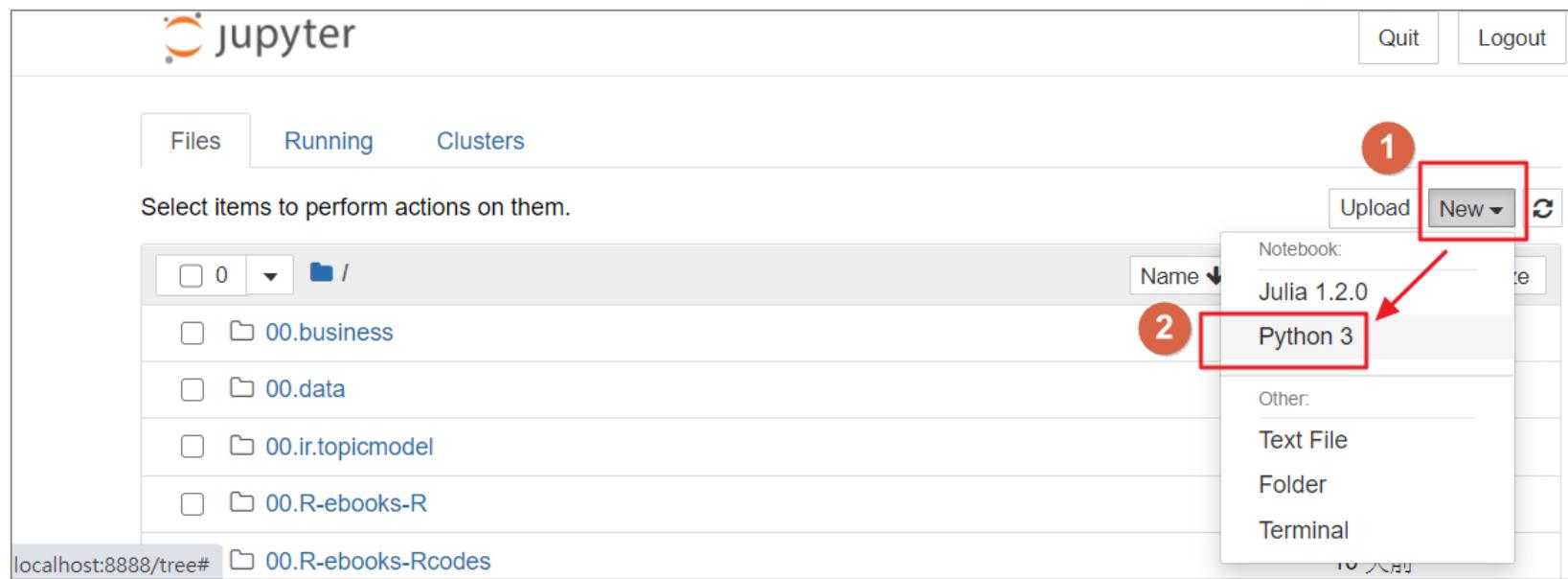
Out[3]: `array([[0.99403652, 0.17109276, 0.84190235, 0.04537563, 0.88868306],
 [0.03845965, 0.8625648 , 0.90504382, 0.71424797, 0.24861374],
 [0.48872416, 0.69003161, 0.51221897, 0.38049437, 0.89148428],
 [0.81011968, 0.30820779, 0.653859 , 0.8477063 , 0.46920767],
 [0.48964363, 0.713421 , 0.65601074, 0.19109199, 0.22990496]])`

In []:

Jupyter-notebook 更改預設目錄

- cd C:\
- jupyter-notebook

或使用 cd \



The screenshot shows the Jupyter Notebook interface. On the left, there's a file tree with several folders like '00.business', '00.data', '00.ir.topicmodel', and '00.R-ebooks-R'. On the right, a context menu is open over a list of options:

- Upload
- New ▾ (highlighted with a red box)
- Notebook:
 - Julia 1.2.0
 - Python 3 (highlighted with a red box)
- Other:
 - Text File
 - Folder
 - Terminal

Two red numbered circles are present: circle 1 points to the 'New ▾' button, and circle 2 points to the 'Python 3' option in the 'Notebook' dropdown.

Jupyter-notebook 快速鍵

- 按 [Esc] cell旁邊為藍色：
 - 按 **x**：刪除當前選擇的cell
 - 按 **a**：在當前選擇的上方新增一個cell
 - 按 **b**：在當前選擇的下方新增一個cell
 - 按 **Shift + Enter**：執行當前的cell並且選到下一個cell
 - 按 **Ctrl + Enter**：執行當前cell
 - 按 **M**：轉成markerdown模式，可以看到紅色框框內容從code變成markerdown



開啟 → Python程式設計-李明昌.ipynb, 瀏覽加入數學式主題

- <http://rwepa.blogspot.com/2020/02/pythonprogramminglee.html>

實作
練習

主題: Python 程式設計-李明昌 - ipynb

檔名: Python_Programming_Lee_ipynb.zip

包括 Python 程式設計-李明昌電子書的原始 ipynb 檔案, 圖檔, 部分資料集

下載: https://github.com/rwepa/DataDemo/blob/master/Python_Programming_Lee_ipynb.zip



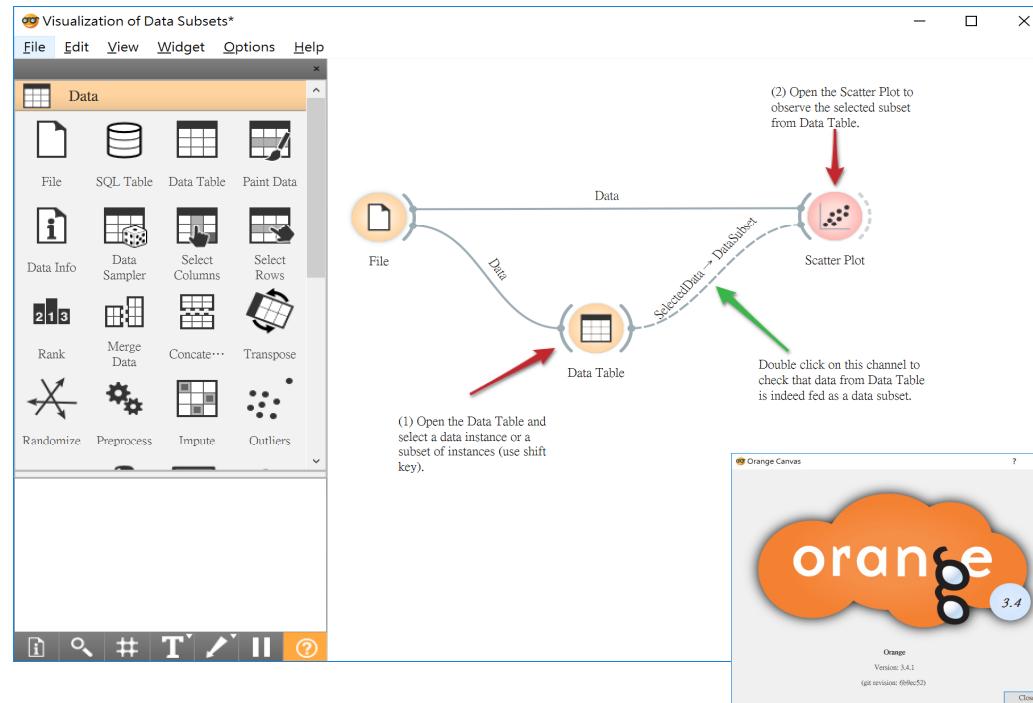
Python_Programming_Lee_ipynb.zip > python.book.lee >	
名稱	類型
.ipynb_checkpoints	檔案資料夾
data	檔案資料夾
img	檔案資料夾
Python程式設計-李明昌.ipynb	IPYNB 檔案

Orange

- YouTube: <https://youtu.be/rh5GxJamtNg>

```
# 安裝 Orange  
# conda install -c conda-forge orange3
```

```
# 使用命令提示列 開啟 Orange  
# python -m Orange.canvas
```

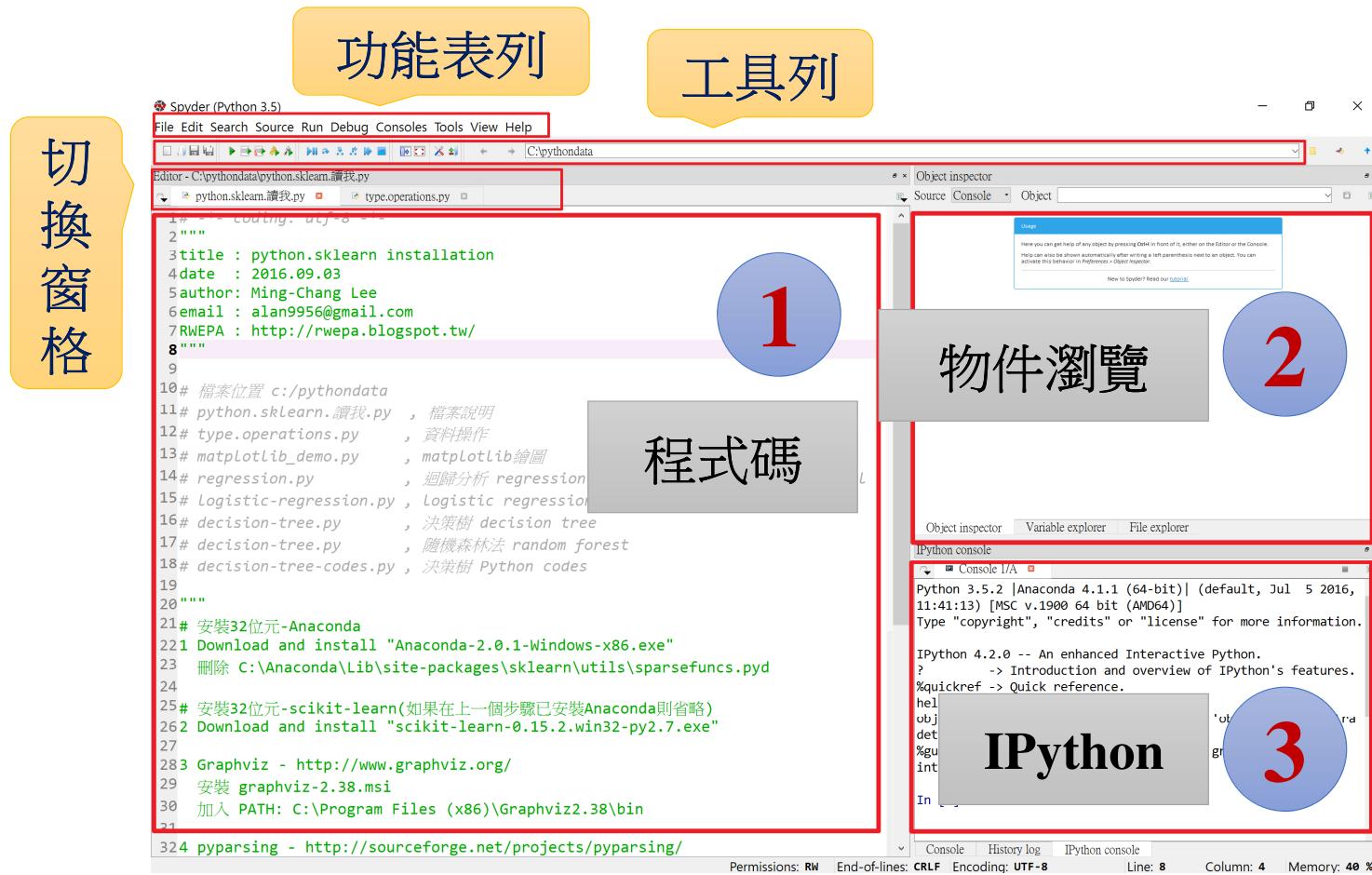


Orange - Scatter Plot



Spyder 軟體簡介

Spyder 畫面



1.7 R,Python專案報告

R 報告

- R + Rmd 檔案
- https://github.com/rwepa/r_data_scientist



R 入門實作班專題報告-日空氣品質指標 (AQI) 之研究

李明昌
2024-09-28

目錄

1. 商業理解	2
2. 資料理解	3
3. 資料準備	11
4. 模式建立 (使用訓練集)	12
5. 評估與測試 (使用測試集)	13
6. 佈署應用與結論	14
參考文獻	15

1

Python 報告

- Python + Jupyter notebook 檔案
- https://github.com/rwepa/python_data_scientist

 tutorial_air_quality_index

Topic: 開放資料 - 空氣品質指標(AQI)(歷史資料)

File: tutorial_air_quality_index.ipynb

Author: Ming-Chang Lee

Date: 2024.11.26

Data source: <https://data.gov.tw/dataset/151824>

References:

https://github.com/rwepa/python_data_scientist

大綱

- 1.商業理解
- 2.資料理解
 - 2.1 模組與資料匯入
 - 2.2 資料摘要
 - 2.3 探索性資料分析(Exploratory Data Analysis, EDA)
 - 2.4 資料視覺化
- 3.資料準備
- 4.建立模型
- 5.模型評估與測試
- 6.佈署應用與結論

參考文獻

參考資料

- RWEPA: <http://rwepa.blogspot.com/>
- RWEPA YouTube: <https://www.youtube.com/@alan9956>
- R入門資料分析與視覺化應用教學(付費):
<https://mastertalks.tw/products/r?ref=MCLEE>
- R商業預測與應用(付費): <https://mastertalks.tw/products/r-2?ref=MCLEE>
- Python 程式設計-李明昌 <免費電子書>:
<http://rwepa.blogspot.com/2020/02/pythonprogramminglee.html>
- 辦公室自動化R + Python 整合應用
<https://quarto.org/>

謝謝您的聆聽

Q & A



李明昌

alan9956@gmail.com

<http://rwepa.blogspot.tw/>