

Power Query 入門

當然，在這個持續更新且永遠在線上的科技時代，點按更新聽起來可能很離奇，但只要做得夠好，則當人們和文化重新創造並更新時，復活就成為可能。

——薩蒂亞·納德拉（微軟公司現任執行長）

本章重點：

- 對 Power Query 有基本的認識並瞭解其起源
- 認識 Power Query 的主要組成部分及 Power Query 編輯器
- 探索此工具並準備範例資料以供分析

在本書中，你將學習如何運用 Power Query 的能力來搞定各種資料挑戰，並在過程中省下多達 80% 的資料準備時間。本章將以正式的介紹起頭。這是 Power Query 應得的。在你閱讀這幾行文字時，可能就有 50 萬名左右的使用者，在這一刻，咬著牙以手動方式在 Excel 中進行重複但卻至關重大的資料處理任務。這種任務他們每天都做，或是每週、每個月都做。

在你讀完本書時，約有 5 千萬人已完成其艱苦的手動資料準備工作，且渾然不覺 Excel 裡就藏著一個工具等著幫助他們簡化該工作。其中有些人已學會使用如 Python 與 R 等進階工具來清理資料，也有些人是一直倚賴其資訊部門，所提出的需求總是要等好幾個月才能獲得滿足，而大部分的人都是為了完成工作，只好認命地花費數百甚至數千小時的時間來準備資料以供分析。若你或你的朋友也是這 5 千萬人中的一份子，那麼是時候該學習 Power Query 並瞭解它將如何改變你們目前所認知的資料分析工作了。

不論你是第一次接觸 Power Query 還是已有使用經驗，本章都將幫助你為前方的旅程做好準備。這段旅程將使你成為一位大師級的資料駕駛者，以及自助的洞見發現者。

Power Query 是什麼玩意兒？

Power Query 是個改變了遊戲規則、令人耳目一新的資料連結暨轉換技術，它存在於 Microsoft Excel、Power BI 等微軟產品中。它讓任何人都能連結豐富的外部資料來源，甚至是試算表中的本機資料，還能利用簡單的使用者介面來收集、合併及轉換資料。一旦資料準備妥當，便可載入至 Excel 及 Power BI 的報表中，又或是以資料表形式儲存在其他包含 Power Query 的產品內。然後，每當該資料有所更動，使用者便能更新其報表以享受資料的自動轉換功效。

另請參閱 自發佈以來，Power Query 已為數百萬人所使用。基於其對資訊工作者及資料分析師能力的重大影響，微軟已決定將之納入更多產品中，包括了：

SQL Server 2017 Analysis Services 與 Azure Analysis Services 的 Microsoft SQL Server Data Tools (SSDT) (詳見 <https://docs.microsoft.com/zh-tw/sql/analysis-services/what-s-new-in-sql-server-analysis-services-2017>)

為 Microsoft PowerApps 所使用的 App 的 Common Data Service (詳見 <https://docs.microsoft.com/zh-tw/powerapps/maker/common-data-service/data-platform-intro>)

當你使用 Power Query 編輯器（亦即 Power Query 的主要介面）來準備資料時，該工具會記錄下一連串的轉換指令，好讓你日後能重複進行該資料準備程序。進階的 Excel 使用者經常將之比作 Excel 中的錄製巨集功能，但其實它比那簡單得多。一旦完成轉換步驟，你就能將準備好的資料載入至報表。而自那一刻起，每當你更新報表，那一連串的轉換步驟便會執行，可省下你數小時甚至數天的重複性手動資料準備工作。

Power Query 真的很簡單易用。它提供了一致的使用者體驗——不論你從哪個來源匯入資料，也不論其格式為何。Power Query 讓你能透過一連串可編輯且易於遵循的小步驟，來實現各種複雜的資料準備方案。在進階的使用情境中，較強的使用者可透過資料編輯列或進階編輯器來修改各個步驟，以自訂轉換程式碼（M 查詢語言的使用將於第九章「Power Query M 公式語言入門」說明）。每個轉換程序都會被儲存為一項查詢，可被載入至報表，或為其他查詢重複使用於建立一系列的轉換元件。

啟。而當你連結如 SQL Server 或 Access 等關聯式資料庫時，其左側窗格還會顯示出更多選項，不過大部分的使用者介面看起來都是一樣的。

如圖 1-9 所示，此交談窗的預設按鈕為「載入」鈕而非「編輯」鈕，因此很多使用者可能會順著預設流程走，於是就完全錯過了 Power Query。按下「載入」鈕你就會跳過資料準備步驟，將該資料表依原樣載入報表中。

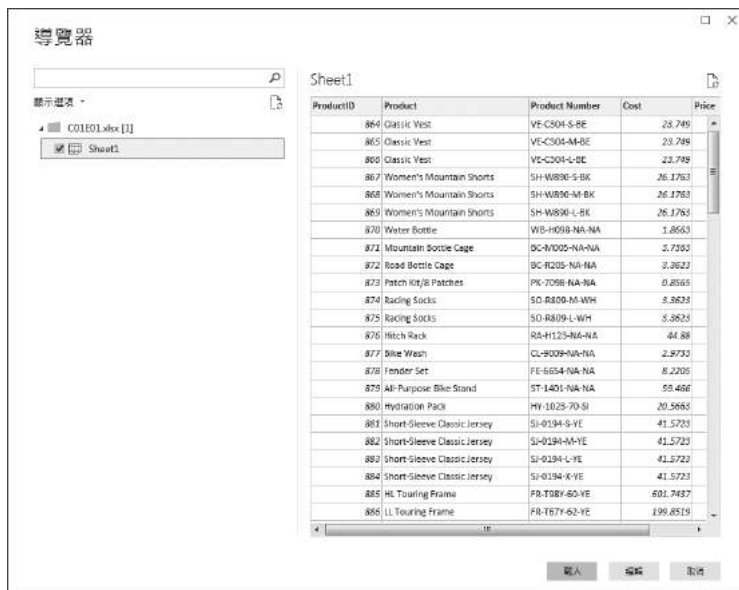


圖 1-9 「導覽器」交談窗讓你選擇要載入的資料表。



Note 若你沒在「導覽器」交談窗（如圖 1-9）中看見「載入」鈕後面接著一個「編輯」鈕，那它可能是以不同的名稱顯示（例如「轉換資料」）。微軟正在評估不同的按鈕名稱，看看哪個名稱最能與客戶產生共鳴。當本書指示說要在「導覽器」交談窗中點按「編輯」鈕，但你找不到時，就請點按「轉換資料」鈕，或是接在「載入」鈕後面的那個按鈕。



Tip 即使你的資料狀況良好，在「導覽器」交談窗中建議你最好還是按「編輯」鈕而非「載入」鈕。這樣就會開啟 Power Query 編輯器，以便你預覽、探索，並確認資料格式符合預期。

練習 2-6：處理來自兩個不同地區的日期

在練習 2-5 中雖有多種日期格式，但我們假設所有日期都來自同一地區。而在此練習中，你將學習如何處理來自不同地區的日期，有些資料列的月和日值應該要交換才行。

請至 <https://aka.ms/DataPwrBIPivot/downloads> 下載活頁簿檔案 C02E06.xlsx，並將之儲存至 C:\Data\C02 資料夾。為了避免過度複雜，此活頁簿只包含取自 AdventureWorks 產品目錄的兩列資料，各列分屬於不同國家的產品，而上市日期的格式也是依據該特定國家之地區設定：

- 國家：美國，上市日期：7/23/2018
- 國家：英國，上市日期：23/7/2018

如圖 2-12 所示，若你 Power Query 編輯器「選項」交談窗中的地區設定是設為「英文（美國）」，則屬於英國（UK）那一列的「Release Date」（上市日期）資料行更改為日期資料類型時，便會傳回錯誤（Error）值；反之若你在該交談窗中的地區設定是設為「英文（英國）」，則屬於美國（US）的那一列便會傳回錯誤值。

上市日期為文字類型

ProductID	Product	Cost	Price	Country	Release Date
864	Classic Vest	23.749	63.5	US	7/23/2018
865	Classic Vest	23.749	63.5	UK	23/7/2018

轉換為日期

地區設定是「英文（英國）」

ProductID	Product	Cost	Price	Country	Release Date
864	Classic Vest	23.749	63.5	US	Error
865	Classic Vest	23.749	63.5	UK	23/7/2018

地區設定是「英文（美國）」

ProductID	Product	Cost	Price	Country	Release Date
864	Classic Vest	23.749	63.5	US	7/23/2018
865	Classic Vest	23.749	63.5	UK	Error

圖 2-12 依活頁簿或檔案的地區設定不同，「Release Date」資料行從「文字」轉換為「日期」時會產生不同的錯誤。

匯總了 Wide World Importers 公司訂單的活頁簿檔案 C02E07.xlsx，可於 <https://aka.ms/DataPwrBIPivot/downloads> 下載。本練習的目標，是要示範使用 Power Query 的「參考」和「移除重複項」命令來將資料表分割為事實表格和查閱表格有多麼容易。

1. 下載活頁簿檔案 C02E07.xlsx，並將其儲存至 C:\Data\C02\ 資料夾。
2. 新增一個空白的 Excel 活頁簿或 Power BI Desktop 報表，然後將 C02E07.xlsx 活頁簿檔匯入至 Power Query 編輯器。
3. 在彈出的「導覽器」交談窗中，選取「Sales_Orders」資料表，再按「編輯」。
4. 將「Sales_Orders」查詢更名為「Sales Orders - Base」。

在此，你的目標是要將原始的資料表分割成兩個具有正確精細度的資料表，如圖 2-14 所示。其中一個資料表（圖 2-14 左下）包含訂單及其庫存商品的識別碼（事實表格），另一個（圖 2-14 右下）則包含庫存商品資料（查閱表格）。

OrderID	Order Date	Quantity	Unit Price	Stock ID	Stock Item	Stock Lead Time
1	1/1/2013	10	230	67	Ride on toy sedan car (Black) 1/12 scale	14
2	1/1/2013	9	13	50	Developer joke mug - old C developers never die (White)	12
2	1/1/2013	9	32	10	USB food flash drive - chocolate bar	14
3	1/1/2013	3	30	114	superhero action jacket (Blue) XXL	12
4	1/1/2013	2	13	50	Developer joke mug - old C developers never die (White)	12
4	1/1/2013	5	32	130	Furry gorilla with big eyes slippers (Black) S	12
4	1/1/2013	96	2.7	206	Permanent marker black 5mm nib (Black) 5mm	14
5	1/1/2013	3	32	121	Dinosaur battery-powered slippers (Green) XL	12

OrderID	Order Date	Quantity	Unit Price	Stock ID
1	1/1/2013	10	230	67
2	1/1/2013	9	13	50
2	1/1/2013	9	32	10
3	1/1/2013	3	30	114
4	1/1/2013	2	13	50
4	1/1/2013	5	32	130
4	1/1/2013	96	2.7	206

Stock ID	Stock Item	Stock Lead Time
1	USB missile launcher (Green)	14
2	USB rocket launcher (Gray)	14
3	Office cube periscope (Black)	14
4	USB food flash drive - sushi roll	14
5	USB food flash drive - hamburger	14
6	USB food flash drive - hot dog	14
7	USB food flash drive - pizza slice	14

圖 2-14 你可將「Sales_Orders」（銷售訂單）資料表分割成事實表格與查閱表格。

5. 在「Sales Orders - Base」查詢上按滑鼠右鍵，選「參考」。然後將新查詢更名為「Stock Items」（要更改查詢的名稱時，你可在「查詢」窗格中，於該查詢上按滑鼠右鍵，選「重新命名」，或直接在「查詢設定」窗格裡「內容」下方的「名稱」欄位輸入新名稱）。

練習 4-6：只轉置資料行名稱

有個效能更佳的做法可將多個資料行名稱轉為單一資料行，以便為轉換表的合併做好基礎準備，那就是只轉置資料行名稱做為過渡階段，之後再套用合併處理。圖 4-8 就說明了你可以如何轉置資料行名稱、套用標準化處理，最後再把這些資料行名稱轉置回原始資料表。

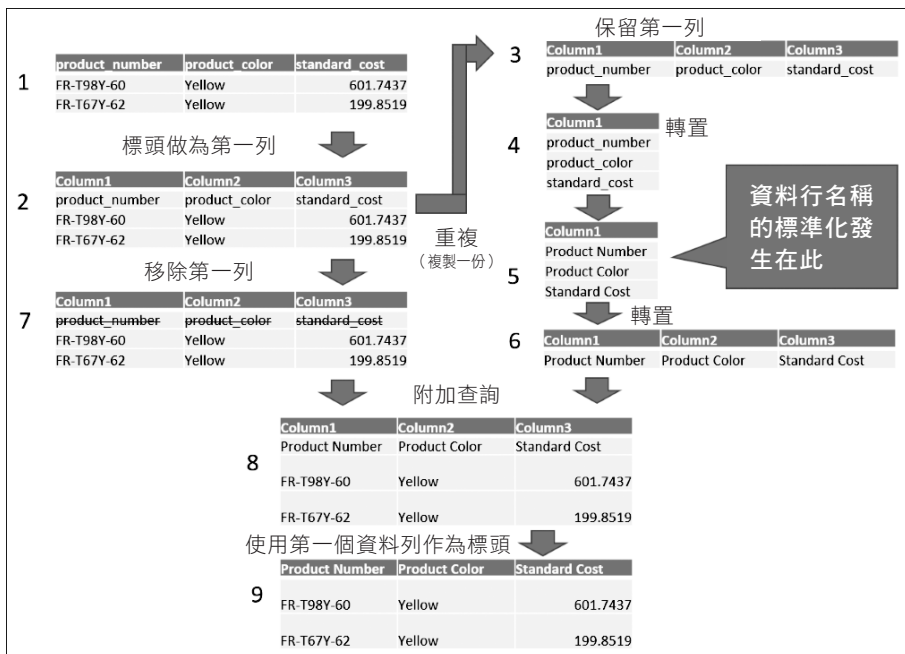


圖 4-8 資料行名稱的標準化包括了九個轉換步驟。

雖然當你以「附加查詢」功能合併幾個特定檔案時，要實行圖 4-8 說明的技巧相當容易，但要將此技巧應用在資料夾時卻會變得相當進階，因為它倚賴自訂函數，故在此建議僅針對學習目的這麼做。若練習 4-4、4-5 所介紹的技巧因效能限制而不適用，然後你又想避免使用練習 4-7 的 M 語言技巧時，你就該運用本例的技巧。

1. 請接續練習 4-4 的步驟 1 ~ 9，從之前存好的活頁簿或 Power BI 報表開始操作。

學習過程中的不同成熟階段

一旦踏上 M 語言的學習旅程，你很可能會經歷如圖 9-1 的六種成熟階段。各階段分別能解決不同類型的資料挑戰，但也可能把你鎖在特定的舒適圈。你願意投注大量精力以進入下一階段嗎？

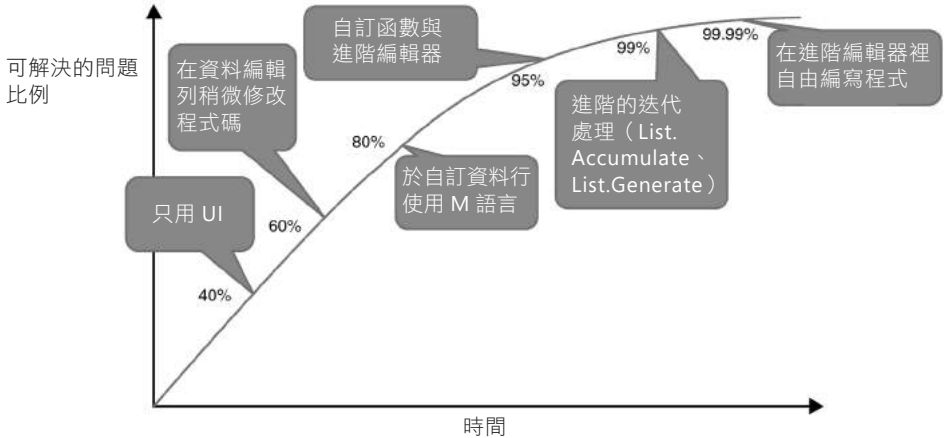


圖 9-1 隨著你逐一達到 M 語言學習過程中的不同成熟階段，你將有能力解決越來越多的問題，不過絕大多數的問題在你到達第二階段時就己能順利解決了。

第一階段：只用 UI

若你是經由本書第一次接觸 M 語言，且已完成前面各章的練習，那你就己到達 M 語言學習過程中的第一個成熟階段，己是個內行的 Power Query 使用者。你知道 M 是存在於背後的程式語言，但可能沒有意願學習它，因為投入的回報並不明確。你也可能害怕學習程式語言，或者看好其他語言甚於 M 語言。同類的程式語言非常多，包括 Excel 公式、VBA、DAX、Python，還有 R 語言 等等，而你可能已經用過其中幾個。那為什麼還要再學一個和你平常用的東西差異很大的 M 語言？

當你發現 UI 只能幫助你解決約莫 40% 的問題時，促使你前往下一階段的動力就會出現。當你變得更積極地想用 Power Query 解決更多資料挑戰時，就會發現很多問題都能靠著簡單的程式碼微調搞定。由於這種微調不需要對語法有深入瞭解，故即使你並不特別精通 Excel 公式，或者根本對軟體程式設計一無所知，這仍是個你會自然到達的階段。

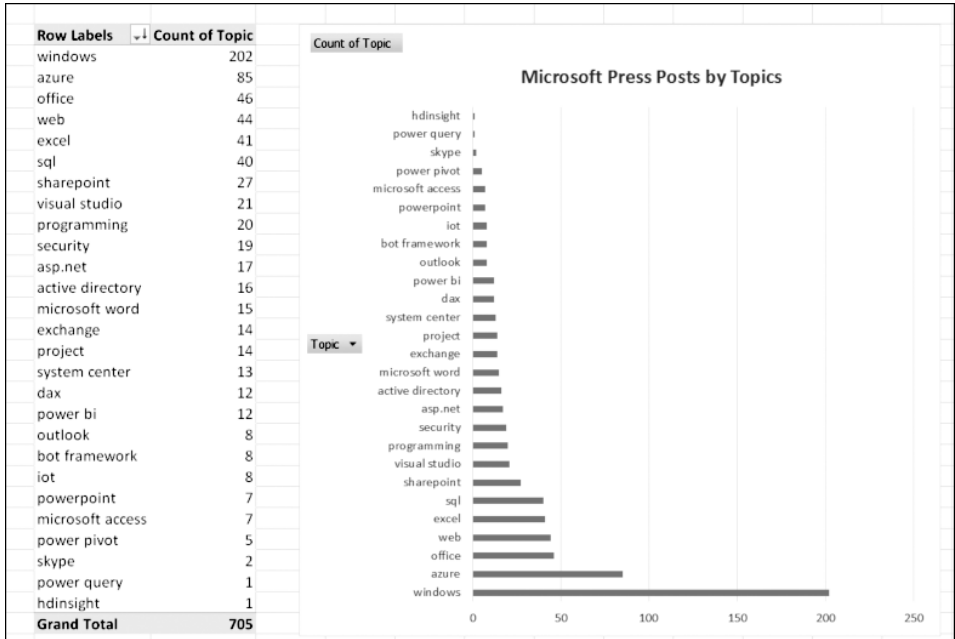


圖 11-16 已完成的解決方案活頁簿顯示出依主題別統計的 Facebook 貼文數量，以此例來說是以 Excel 的樞紐分析表及樞紐分析圖來呈現。

練習 11-7：在 Power BI 中建立文字雲

在本章最後，該是時候看看 Power BI 如何能以文字雲來視覺化文字資料行了。在前一節中，你已學到如何分割並計算字彙數量，以及在報表中忽略停用字。但事實上，以小型資料集來說，你可將所有訊息直接載入至 Power BI 的文字雲，完全不需分割字彙。

本練習將帶領你體驗文字雲自訂視覺效果的設定，並示範如何忽略停用字而不需分割字彙。只要依以下步驟操作，你就能做出如圖 11-17 的文字雲。

1. 啟動 Power BI Desktop。
2. 在「常用」索引標籤的「自訂視覺效果」功能區，點按「從 Marketplace」。
3. 待「Power BI 自訂視覺效果」交談窗開啟，於搜尋欄位輸入「Word Cloud」後，按下搜尋圖示。

通常，在進階的文本分析實作中，你會有個複雜的架構圖，其中有各式各樣的元件，各自專注於特定複雜任務組中的許多元素。幸好，微軟已將所有複雜的工作都實作於 Azure 認知服務中，你可把它看成是個神奇的黑盒子。只要將文字傳入那個黑盒子，它就會傳回已翻譯成你想要語言的文字、情緒評分，或是抽出的關鍵片語清單。因此你的架構就變得相當基本（只有一個 Excel 活頁簿或 Power BI 報表），其中所有的資料準備與資料移動都由 Power Query 引擎完成，然後由黑盒子發揮神奇魔法，傳回富含意義的文字。

圖 12-2 說明了此解決方案中的三個主要步驟，你實作一個從外部資料來源載入文字資料行的報表，然後將文字資料餵給 Azure 認知服務的 Translator Text API 或 Text Analytics API，最後再由 Azure 認知服務將結果傳回給報表。必須注意的是，這三個步驟是在單一更新操作中進行。當你完成查詢後，每次更新報表，這三個步驟就會再次實行，並提供所需結果。

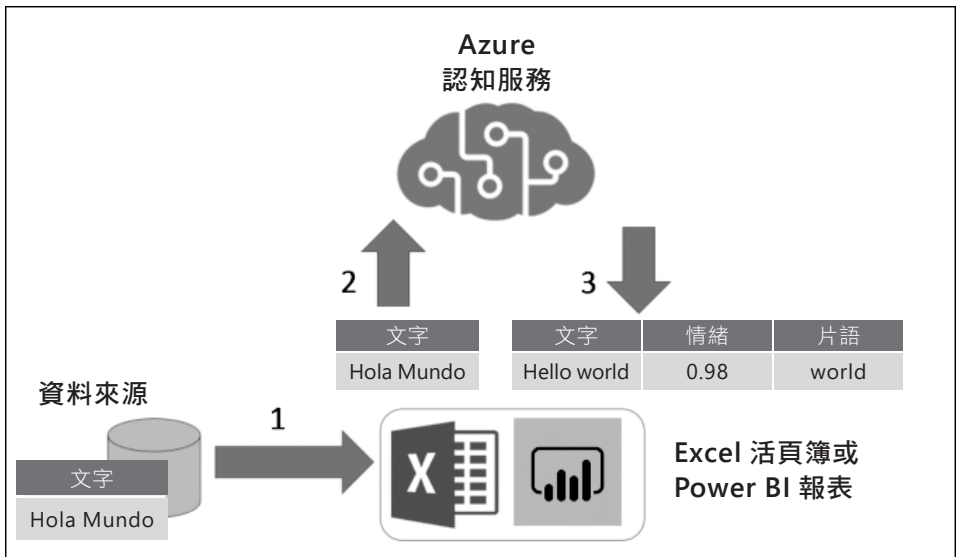


圖 12-2 一次更新需要進行三個步驟，從外部資料來源到認知結果。

Azure 認知服務是個「軟體即服務」(SaaS, Software as a Service) 的例子。這表示你不需撰寫任何伺服器端的程式碼，也不必擔心執行此工作的外部伺服器或電腦的可用性。你只需要連接資料集、準備好資料、上傳至 Azure 認知服務，然後獲得回應並將之載入至報表即可。這些全都在單一報表中完成，而且是在 Power Query 的大力協助下。報表與認知服務間的介面，是採用

即使無法進一步遍歷超過第二層級朋友的圖譜，你依舊能在 Power BI 中建構出如圖 13-4 所展示的社交圖譜。

你可下載已完成的解決方案檔 C13E02 - Solution.pbit（如圖 13-4 所示），或是查看 C13E02 - Solution.xlsx 檔中的查詢。這兩個檔案都可至 <https://aka.ms/DataPwrBIPivot/downloads> 下載。



Tip 透過進階的 API 叫用，你能夠更有效率地遍歷社交圖譜。例如，以下的 M 函數做為單一行程碼使用時，會叫出你的 100 個朋友，以及這些朋友的前 50 個朋友：

```
=Facebook.Graph("https://graph.facebook.com/v3.0/me?fields=friends.limit(100){name,id,friends.limit(50){name,id}}")
```

練習 13-3：找出你朋友按過讚的頁面

這段練習是練習 13-2 的延伸，在此你不匯入朋友的朋友，而是要匯入你朋友曾經讚過的頁面。延續前一段練習的限制，在這裡你同樣只能匯入一部分的朋友——你只能看到那些曾透過 Power Query 匯入 Facebook 資料的朋友。但你將能瞭解他們喜歡哪些頁面，以及由其按讚頁面所反映出的共通興趣為何，就如圖 13-5 所示。

1. 新增一個空白的 Excel 活頁簿或 Power BI Desktop 報表。

在 Excel 中：於「資料」索引標籤選取「取得資料 > 從線上服務 > 從 Facebook」。

在 Power BI Desktop 中：於「常用」索引標籤點按「取得資料」鈕。待「取得資料」交談窗開啟，點選「線上服務」分類後往下捲動，點選「Facebook」，再按「連接」鈕。你也可在「搜尋」欄位輸入「Facebook」後點選「Facebook」項目，再按「連接」。

2. 在「Facebook」交談窗中，於第一個文字欄位保留「me」，再將「連接」選單選為「朋友」後，按「確定」鈕。當你看到一交談窗顯示出資料內容的預覽時，請按下「編輯」鈕。

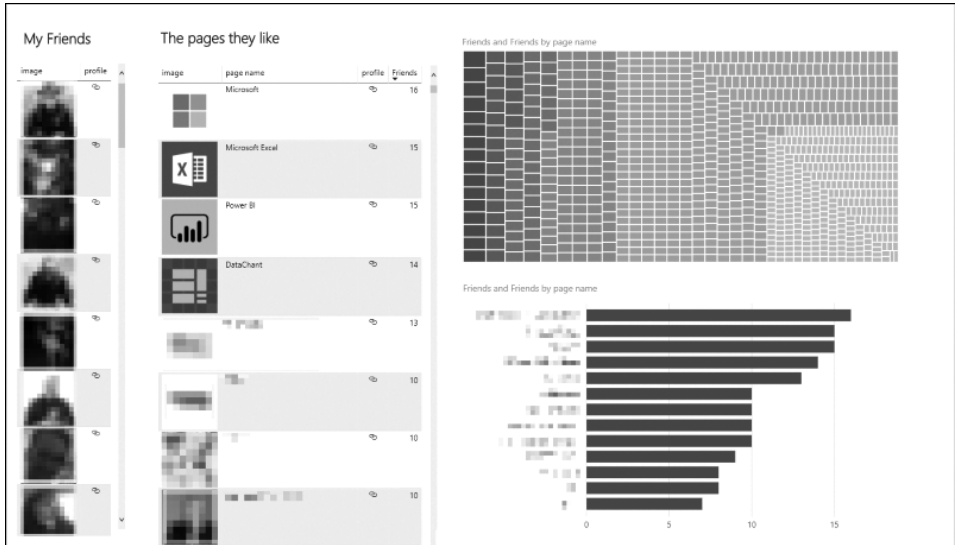


圖 13-5 你可用 Facebook 的大頭貼照與超連結來豐富你的報表，詳情請見解決方案檔 C13E03 - Solution.pbix。

3. 為了找出你的朋友們曾對哪些頁面按讚，請展開「object_link」資料行。在「展開」窗格取消所有除了「connections」以外的資料行，並確定「使用原始資料行名稱做為前置詞」核取方塊已取消，再按「確定」鈕。接著展開「connections」資料行。在「展開」窗格取消所有除了「likes」以外的資料行，並確定「使用原始資料行名稱做為前置詞」核取方塊已取消，再按「確定」鈕。
4. 將目前的查詢更名為「Base」，然後於「查詢」窗格中的「Base」查詢上按滑鼠右鍵，選「參考」。
5. 將新查詢更名為「Friends」，並趁此「Friends」查詢處於選取狀態時，刪除「likes」資料行。
6. 至「查詢」窗格，在「Base」查詢上按滑鼠右鍵，選「參考」。將新查詢更名為「Pages」，並趁此「Pages」查詢處於選取狀態時，刪除「name」與「id」資料行。
7. 繼續在選取「Pages」查詢的狀態下，展開「likes」資料行。請在「展開」窗格取消所有除了「name」和「id」以外的資料行，並確定「使用原始資料行名稱做為前置詞」核取方塊已取消，再按「確定」鈕。

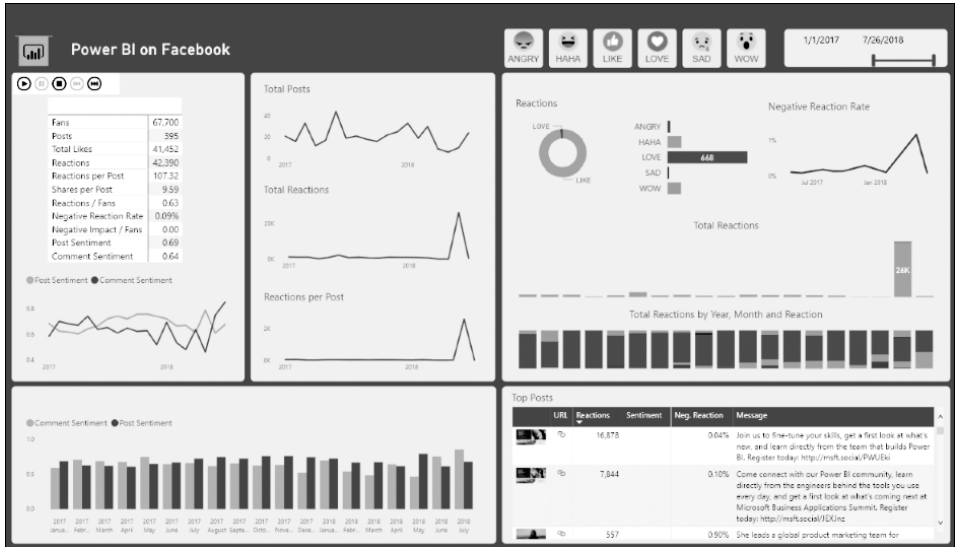


圖 13-6 你可用 Power Query 分析任意 Facebook 頁面，並以 Text Analytics（在第十二章學到的）取得貼文與留言的情緒評分。

練習 13-4：從 Facebook 頁面抽取貼文及留言 —— 基本做法

在第十一章「基礎文本分析」中，你曾分析微軟出版社的 Facebook 頁面貼文。而在本練習中，你將學習從任意 Facebook 頁面匯入貼文與留言的基本步驟。但就如你將學到的，這些步驟對於分析所有頁面來說很沒效率，且在載入留言數眾多的頁面時可能會太過緩慢。

1. 新增一個空白的 Excel 活頁簿或 Power BI Desktop 報表。

在 Excel 中：於「資料」索引標籤選取「取得資料 > 從線上服務 > 從 Facebook」。

在 Power BI Desktop 中：於「常用」索引標籤點按「取得資料」鈕。待「取得資料」交談窗開啟，點選「線上服務」分類後往下捲動，點選「Facebook」，再按「連接」鈕。你也可在「搜尋」欄位輸入「Facebook」後點選「Facebook」項目，再按「連接」。