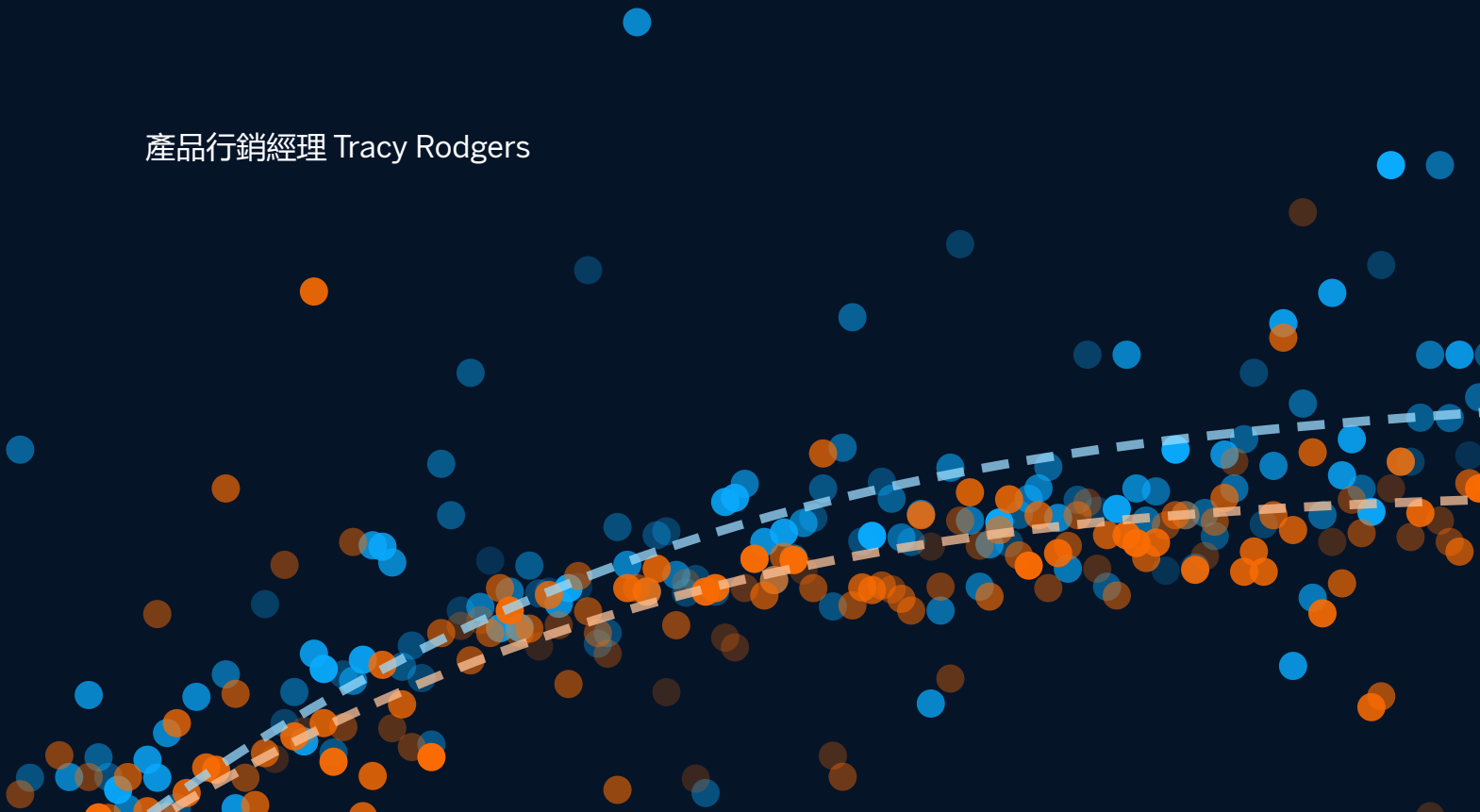




運用 Tableau 進行進階分析

產品行銷經理 Tracy Rodgers



引言

我們的世界，原本界線分明。除非瞭解如何進程式設計，否則進階分析技法便遙不可及。需要知道如何使用 R、Python 及/或 SAS，不然就得要將主要分析工作交由他人處理。在 Tableau，我們相信若要真正運用科技擴增人類智慧，便需要為各種技能水平的使用者提供豐富的功能，其中包括進階分析功能。

Tableau 的分析深度能為 Netflix 與 Spotify 等頂尖分析企業的資料科學群組增強工作流程。只要點幾下滑鼠，就能建立盒狀圖、樹狀圖、預測性視覺化呈現，以及同類群組分析。您可以連接 R、MATLAB 和 Python，並使用 Tableau 來擴充分析功能，將模型結果視覺化。無技術背景的使用者能提出之前無法處理的問題，資料科學家則能透過迭代，更快發現更深入的分析見解，找出更有用的發現。

在本白皮書中，我們將探討 Tableau 能如何在分析專案的各個階段提供協助，不過會將焦點特別放在幾項進階功能上。

我們將會說明以下情境，以及支援這些情境的功能：

細分和同類群組分析： Tableau 有助於產生直觀的調查性流程，進而促進快速、彈性的同類群組分析。您可以從任意數量的維度切分資料，在不同的觀點間快速進行切換。如要使用機器學習擴增細分分析作業，自動化聚類分析可協助您在維度眾多的大型資料集中找出難以發現的模式。

情境和假設分析： 藉由結合 Tableau 的彈性前端和強大輸入功能，您可以快速修改計算並測試不同的情境。

複雜計算與統計功能： 有了 Tableau 的強大計算語言，您將能透過任意計算擴增分析內容，並且使用簡潔的運算式執行複雜的資料操作。從巢狀放置基本彙總，到標準差、眾數、峰度、偏度、相關性與共變異數等統計計算，Tableau 都能執行。

時間序列與預測性分析： 由於世界上許多資料都能建立成時間序列的形式，Tableau 原生支援豐富的時間序列分析功能，並提供完善的 UI 供您探索季節性與趨勢、進行資料取樣、執行預測等預測性分析及其他常見的時間序列作業。

外部服務整合： R、MATLAB 與 Python 整合功能不僅提供 Tableau 前端的強大功能與易用性，也能讓專家運用先前在其他平台上完成的工作，並處理細密複雜的統計與機器學習相關需求。

目錄

細分和同類群組分析	4
聚類分析.....	5
集合與集合動作.....	6
群組.....	7
假設與情境分析	9
參數.....	9
故事點.....	10
複雜計算	10
詳細等級運算式.....	10
表格計算.....	11
時間序列與預測性分析	12
時間序列分析.....	12
預測.....	14
外部服務整合	16
Python、R 與 MATLAB 整合功能.....	17
總結	19
關於 Tableau	20
相關白皮書	20
瀏覽其他資源	20

1. 細分和同類群組分析

功能：聚類分析、集合與集合動作、群組

為了產生初始假說，商務使用者與資料專家通常都是以相同的方式開始著手，即建立細分區段並/或執行非正式的同類群組分析。

對不同的細分區段提出一系列的基本問題，能協助分析師瞭解資料，並驗證假說，像是「使用信用卡付款的客戶，是否比使用支票付款的客戶更容易留住？」具有快速進行迭代的能力，有助於促進模型開發，確保專案進度不拖延。

Tableau 備有豐富的功能組合，能快速進行迭代分析，並比較細分區段。舉例來說，在 Tableau 中，只需要幾個計算欄位，稍微拖放一下，就能建立一個按照星期幾與時間呈現病患人數的儀表板，還能在儀表板上根據部門瀏覽趨勢（圖 1）。

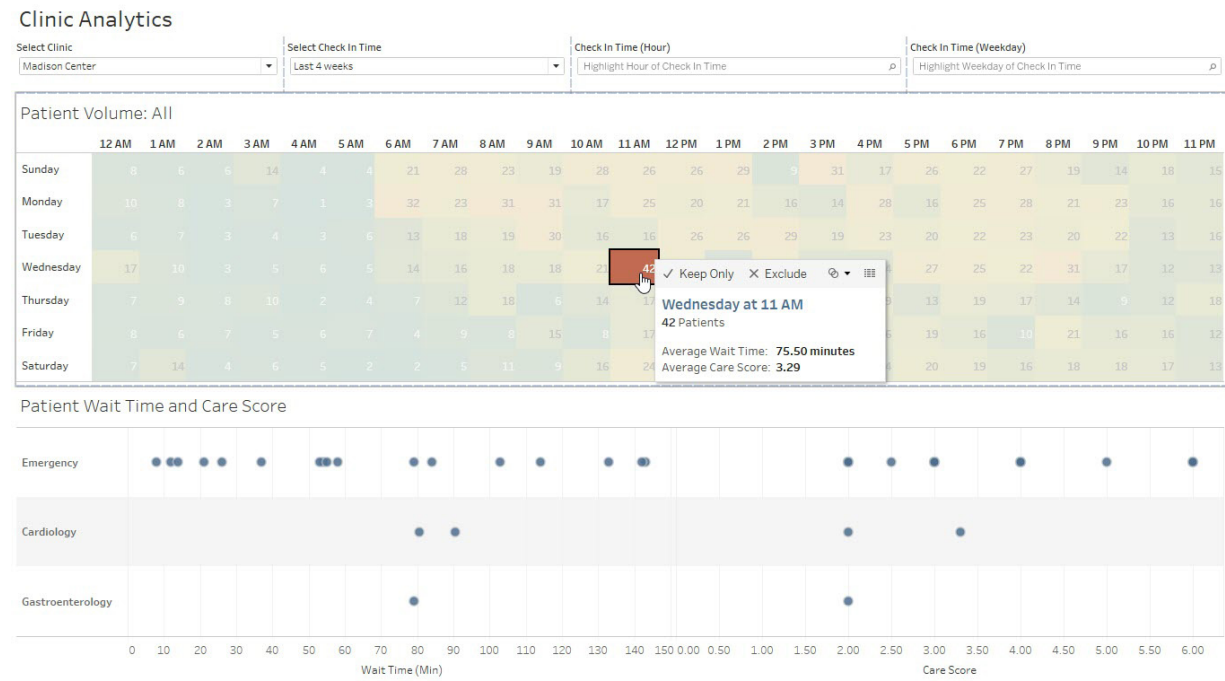


圖 1：這個互動式診所儀表板顯示一週中特定一天與時間的病患人數，就診的原因，以及等待時間。

聚類分析

Tableau 中的聚類分析是另一個使用非監督式機器學習技術來細分資料的選項，非常適合在需要考量大量變數時使用。

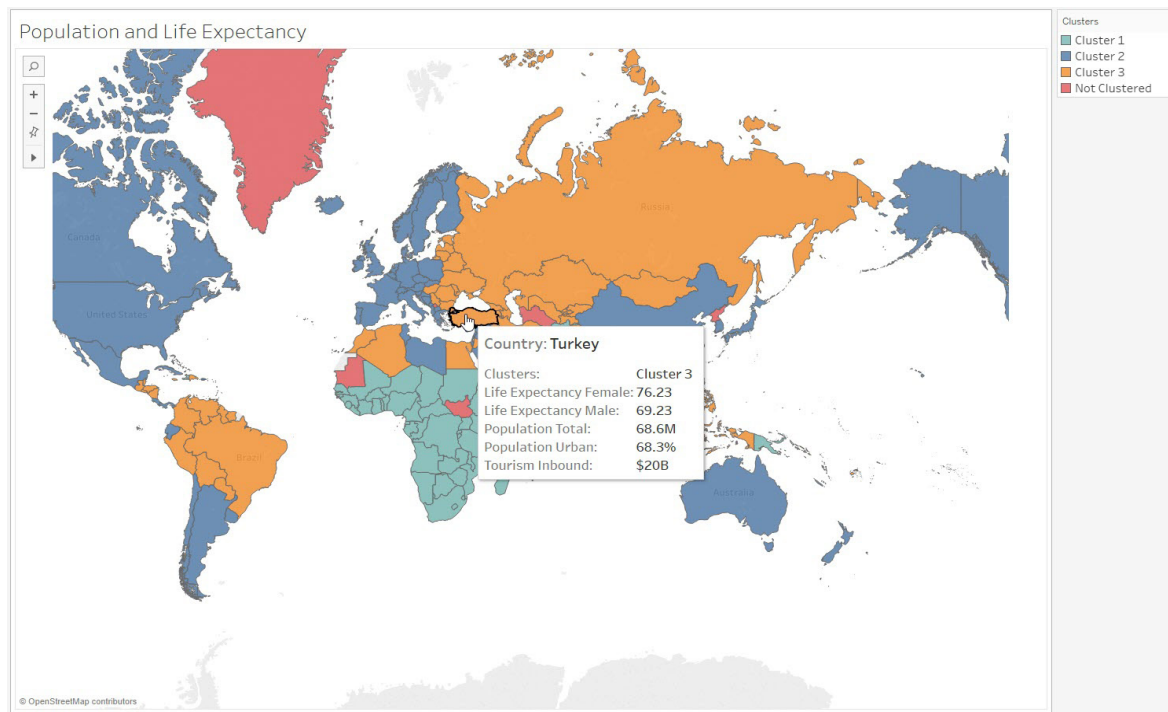


圖 2：使用聚類分析自動細分資料。

圖 2 中可看到依據聯合國開發程度指標，對國家或地區進行聚類分析的結果。演算法建議使用三叢集解決方案，讓低度開發、開發中與已開發國家或地區各有一個對應的聚類。

透過 Tableau 靈活彈性的介面，您也能輕鬆地測試不同的理論，並探索各同類群組的分布。能夠透過視覺方式進行迭代，即可省下在其他工具中調整指令碼並重新執行模擬的時間。

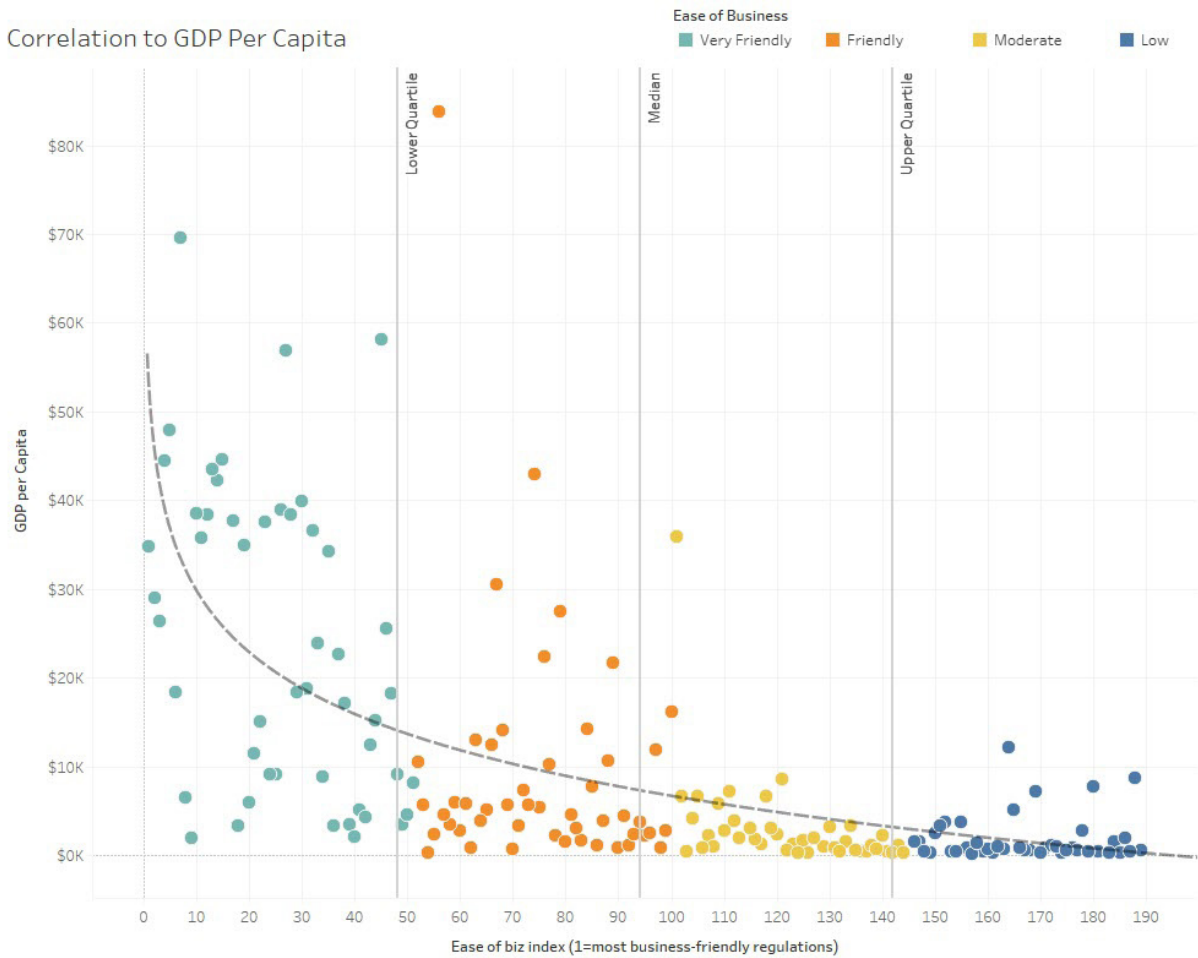


圖 3：在幾秒鐘內細分並瀏覽資料。此範例按「法規企業友善程度」和人均 GDP 細分國家或地區。

如圖 3 所示，只要將聚類分析使用的細分欄位和度量拖曳到畫布上，便能突顯四個國家或地區聚類之間的差異，產生出新的資料檢視方式。

集合與集合動作

使用**集合**即可透過手動選取的方式或使用程式設計邏輯來定義資料物件的集合。集合在許多情境中都非常實用，例如篩選、突出顯示、同類群組計算，以及異常值分析。您也能將多個集合加以結合，藉此測試不同的情境，或建立多個同類群組進行模擬，例如結合獨立產生的數個不同客戶群組來進行留客率分析，或套用多個連續的條件。

提供**集合動作**功能，表示您可將可視化項中的一組所選資料點儲存在集合中。任何工作表中的使用者選取動作，都會以單一、經過協調的選取動作更新集合，並修改所有的目標工作表。如此一來，即可進行如比例筆刷等多種使用案例（圖 4）。

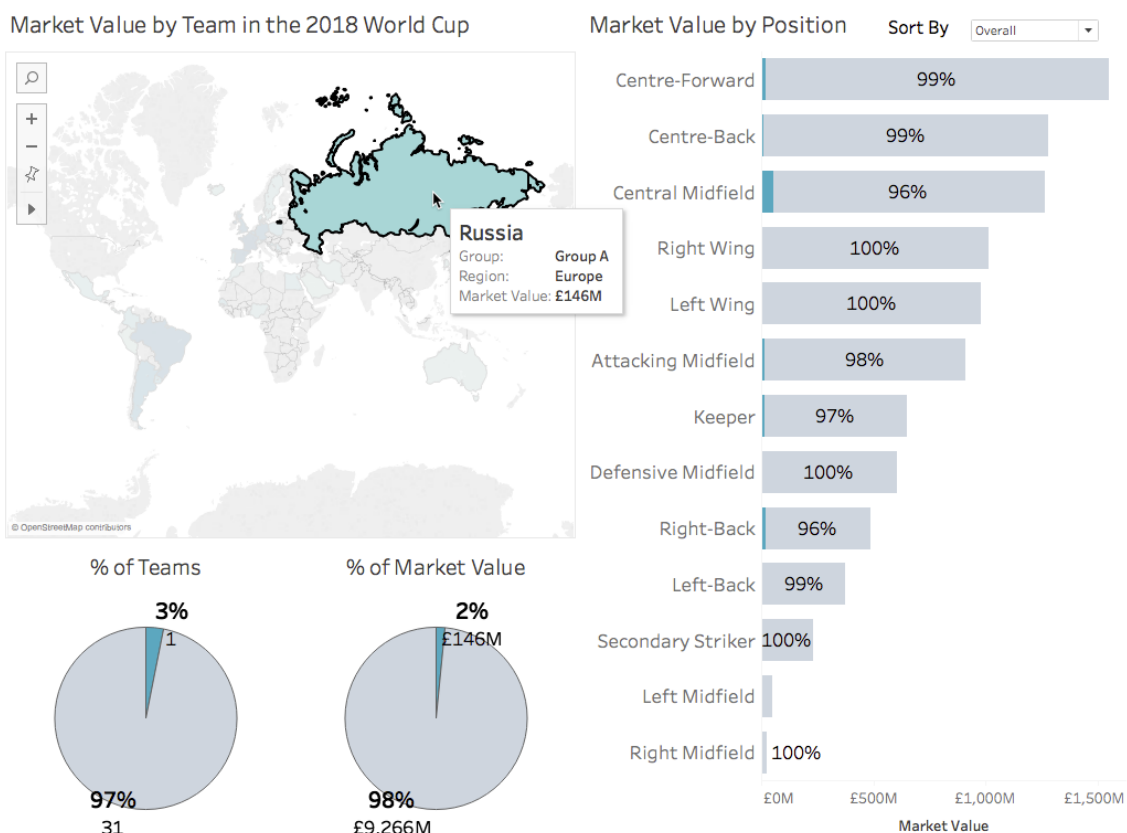


圖 4：集合動作讓使用者能夠進行**比例筆刷**操作。比例筆刷是一種互動式分析技法，能以總規模為基準顯示所選部分的相對規模。選取「Russia」（俄羅斯）即可按世界盃球員位置顯示該國的市場價值。

群組

為滿足建立臨時類別與層次結構的需求，Tableau 提供一個稱為**群組**的功能。對於基本資料清理方面的需求，群組功能也能提供協助。

透過群組，使用者能根據要進行的分析工作，將相關的成員結合在一個欄位中，為資料建立直觀的結構。舉例來說，如果您正在處理按主修科系顯示平均測驗分數的視圖，您可能想要將特定主修科系分為同一群組，藉此建立不同的類別。英文和歷史可能會組成「人文主修」群組，而生物學和物理學則可能會被分到「科學主修」群組。這樣的安排可讓分析師在分析的過程中自訂呈現方式，並控制資料的彙總作業。

資料有一致性與品質問題時，群組功能也能提供一些助益。舉例來說，加利福尼亞州可能會以全名稱呼，不過也有可能以縮寫 CA 或「加州」進行稱呼。分析師與商務使用者通常沒有權限可透過直接變更來源系統來解決這類問題。這表示微小的資料錯誤，便可能對探索性分析造成極大的阻礙。需要停止提問才能要求資料變更，不僅導致專案延遲，也會打斷想法的快速發展。透過群組功能，您將能快速定義新的細分區段，並使該區段包含分析中所用的所有替代名稱，如此一來即可繼續提問，不必中斷流程。

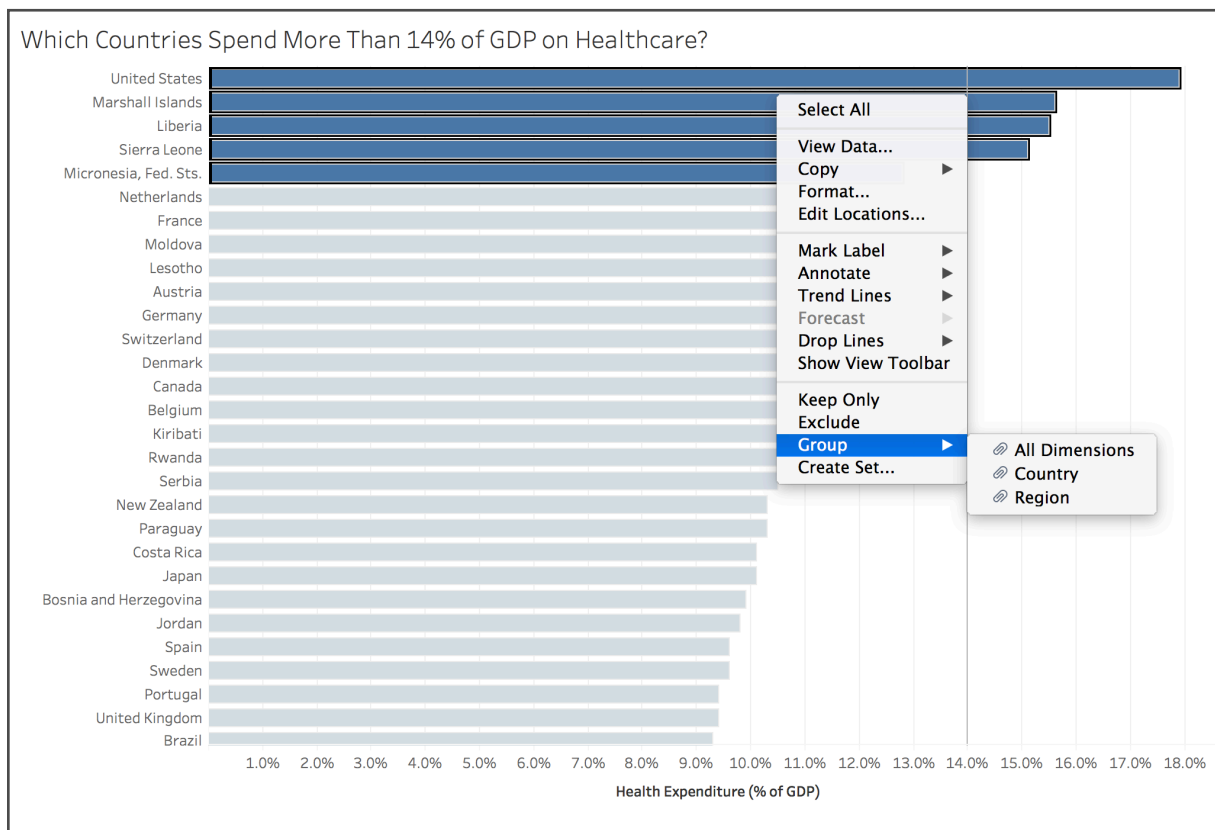


圖 5：透過建立群組，在一個欄位中結合相關的成員。在此範例中，您可能會想要將醫療保健花費占 GDP 最大比例的前五個國家或地區分成一個群組。

此外，上述所有功能本身皆具備簡易的更新機制。如果在 Tableau 中選擇即時連線並更新資料，分析與分析底下的所有元件（如集合和群組）也會一併更新。這表示您不必手動重新執行報告或相依指令碼，同類群組成員組成也會自動更新。簡易更新機制能協助減輕報告作業的負擔，也是另一種測試情境的方法。簡易更新機制可透過抽換基礎資料，在不更新分析堆疊的情況下探測分析對於初始條件的敏感度。

Tableau 為使用者提供快速細分資料並將資料分類的功能，讓商務使用者與分析師能輕鬆地執行強大的同類群組分析。這些功能可協助資料科學家查驗初始假說，並測試各種情境。

2. 假設與情境分析

功能：參數、故事點

在某些情況下，使用者可能會想要測試變更某個或某組特定的值，會如何影響分析的輸出結果。此類操作可用於測試不同的理論、為同事加強識別重要的情境，或調查新的業務可能性。透過 Tableau，您能在分析中嘗試使用不同的輸入。Tableau 中提供簡易的控制項，可讓您輕鬆地標記與分享各個情境，同時讓資料保持最新狀態。

參數

執行假設分析時，您可能會想要變更計算的基值，重新定義配額，或設定初始條件。運用 Tableau 中的參數即可輕鬆完成這些工作。定義參數，是一種變更模型或儀表板輸入值的方法。參數可驅動計算、變更篩選器閾值，甚至能選取哪些資料會送入儀表板。無技術背景的使用者可利用參數來嘗試不同的輸入，探索複雜模型可能產生的輸出。

除了協助您測試假設，Tableau 的參數功能也能讓您在互動式報告中呈現假設分析的結果。圖 6 中的銷售佣金假設分析是由參數所驅動。銷售經理可嘗試不同的佣金比例、底薪和配額，且能透過即時回饋資訊得知這些變更如何影響關鍵指標。

Sales Commission Model

Enter new quota, commission rate and base salary to estimate sales and compensation

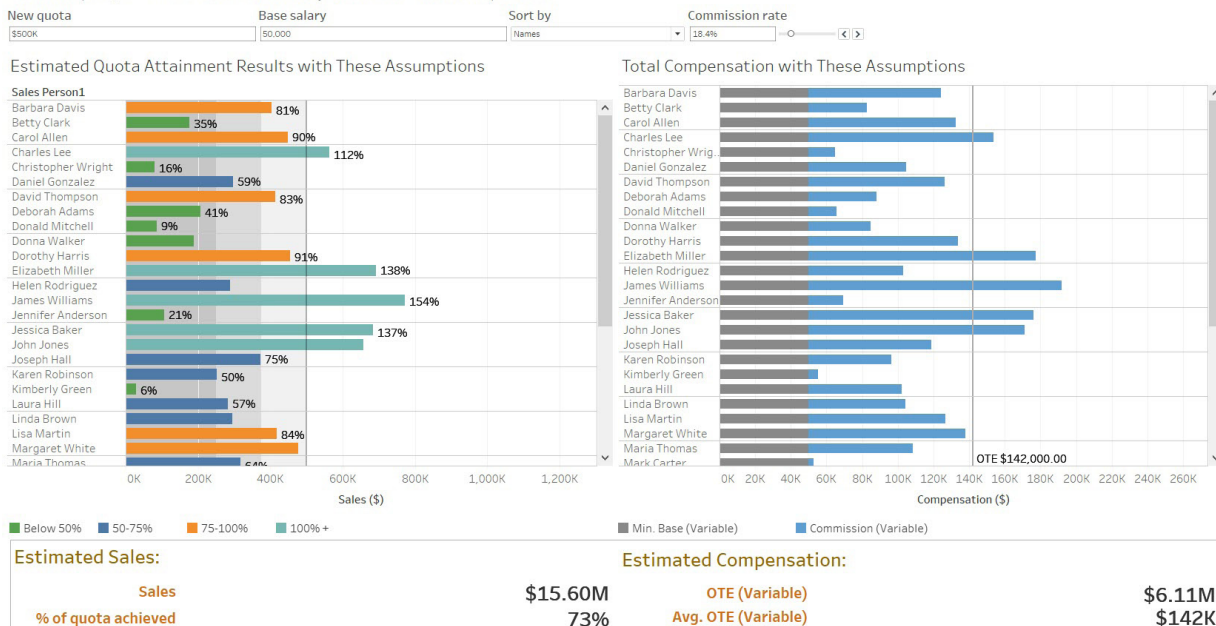


圖 6：在這份參數驅動的銷售報告中，操控者可探索配額、佣金與薪資在組織中的影響。

故事點

故事點是 Tableau 使用資料建立脈絡的方式，若與參數搭配使用，即可為有趣的結果擷取快照，然後繼續進行探索。透過故事點，您將能建構隨著資料變更及可視化項修改而更新的簡報。不過故事點非常聰明，會保留參數值，因此使用者可放心地標記並返回特定情境，不需要中斷分析流程才能返回。您也可以比較數組不同輸入的結果，不必擔心螢幕擷取畫面並非最新，或是需要重新執行模擬。

Tableau 透過集合、群組、拖放細分及參數，讓使用者能夠從理論和問題進展到使用專業的儀表板提出問題並自行測試各種情境，而且即使不是專家也能順利提問與測試。簡化假設分析，能讓資料相關專業人士將心力專注在分析中較複雜的層面上，並提供更強大的分析見解，讓最終使用者也能透過直觀易懂的視覺化呈現與資料互動。如此的參與度提升，有助於組織上下推動變革，做出更明智的決策。

3. 複雜計算

功能：計算欄位、詳細等級運算式、表格計算

一般來說，來源資料並未包含進行全面性分析所需的所有欄位。分析師需要使用簡單但功能強大的語言來轉換資料，並定義精密複雜的邏輯。

Tableau 雖然簡單易用，卻也提供支援複雜邏輯的強大功能。Tableau 中的**計算欄位**可供您根據資料來源中已存在的資料建立新資料。您可以使用計算欄位輕鬆執行算術運算、運算式條件邏輯，或執行特定資料類型的特殊運算。提供進階分析的兩種計算欄位為詳細等級運算式和表格計算。

詳細等級運算式

詳細等級運算式為 Tableau 計算語言的延伸，功能強大，能夠在單一可視化項中回答牽扯多個細微度等級的問題。使用詳細等級運算式，即可以簡潔的運算式處理先前無法或難以解決的情境。

詳細等級運算式能大幅簡化同類群組分析（如前一部分所述）與多次執行形式的彙總。圖 7 按區域市場顯示新獲客戶數的累積加總，並按第一筆訂單的日期將客戶分為「新客戶」和「現有客戶」兩個同類群組。我們會在下一個部分說明時間序列分析，也會探討實現這項分析的其他計算語言層面。從這張圖表可以看出從一月開始到該年度結束，北美洲的新獲客戶數呈現下降狀態。詳細等級運算式將原本需要複雜 SQL 分組依據陳述式的細分作業，變成可在 Tableau 前端編輯和調整的簡單、直觀運算式。

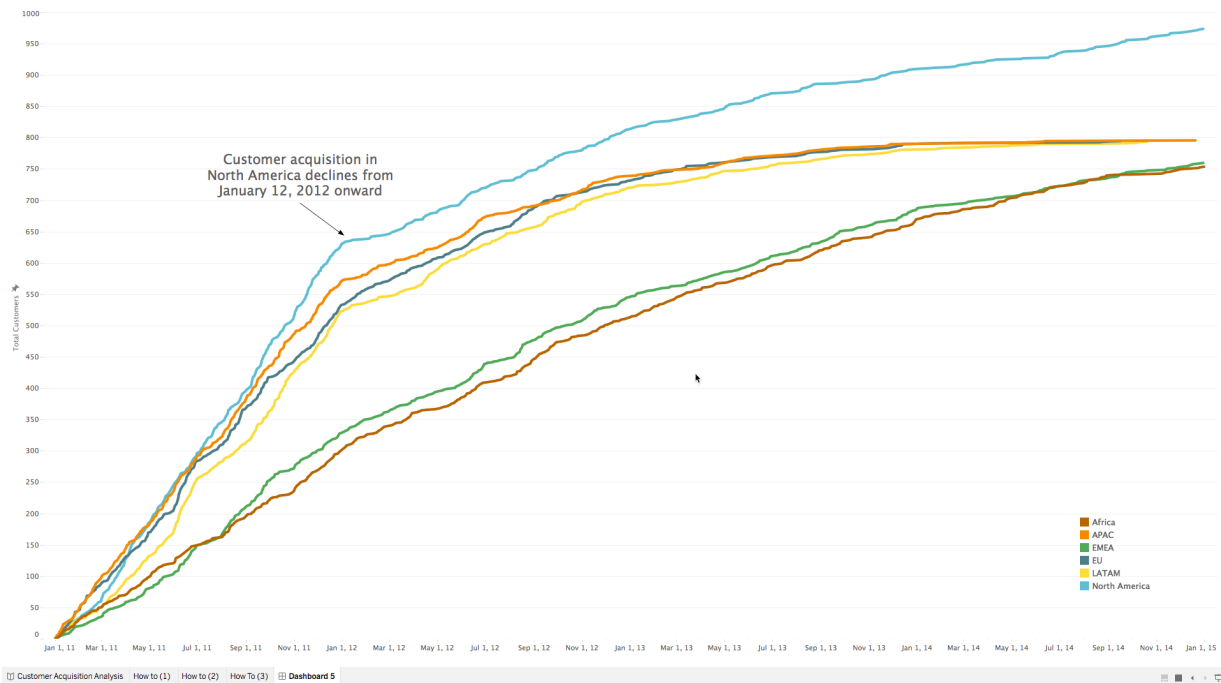


圖 7：這個可視化項使用詳細等級運算式按市場呈現每日的新獲客戶總人數趨勢。使用詳細等級運算式能確保回訪客戶不會被設計為新客戶。

表格計算

Tableau 的**表格計算**功能可讓使用者執行本質上屬於相對性質的運算。更具體來說，表格運算是套用到表格中所有值的一種運算，通常仰賴資料表結構本身。這種類型的計算包含許多時間序列運算，例如滯後時差或累積加總，不過也提供排名與加權平均等計算。

在 Tableau，有兩種方式可以使用表格計算。第一種方式是透過一組常用的表格計算，這組計算稱為**快速表格計算**。這些計算提供一鍵定義表格計算的功能，是絕佳的起點。您也可以使用**表格計算函數**自行建立表格計算。透過這些函數，活頁簿作者便能精準操控結果集合。此外，由於所有表格計算都是以相同的計算語言表示，您可以從其中一個快速表格計算開始操作，如果需要更複雜的計算，只要手動編輯即可。

運用表格計算，只要按幾下滑鼠，或使用簡單的運算式，就能解決具挑戰性的資料庫工作，例如操控彙總資料，以及建立仰賴資料架構的彙總等。表格計算不僅能為無技術背景的使用者提供許多助益，也能為專家省下大量的工時，並免除撰寫 SQL 程式碼這項勞心費神的工作。

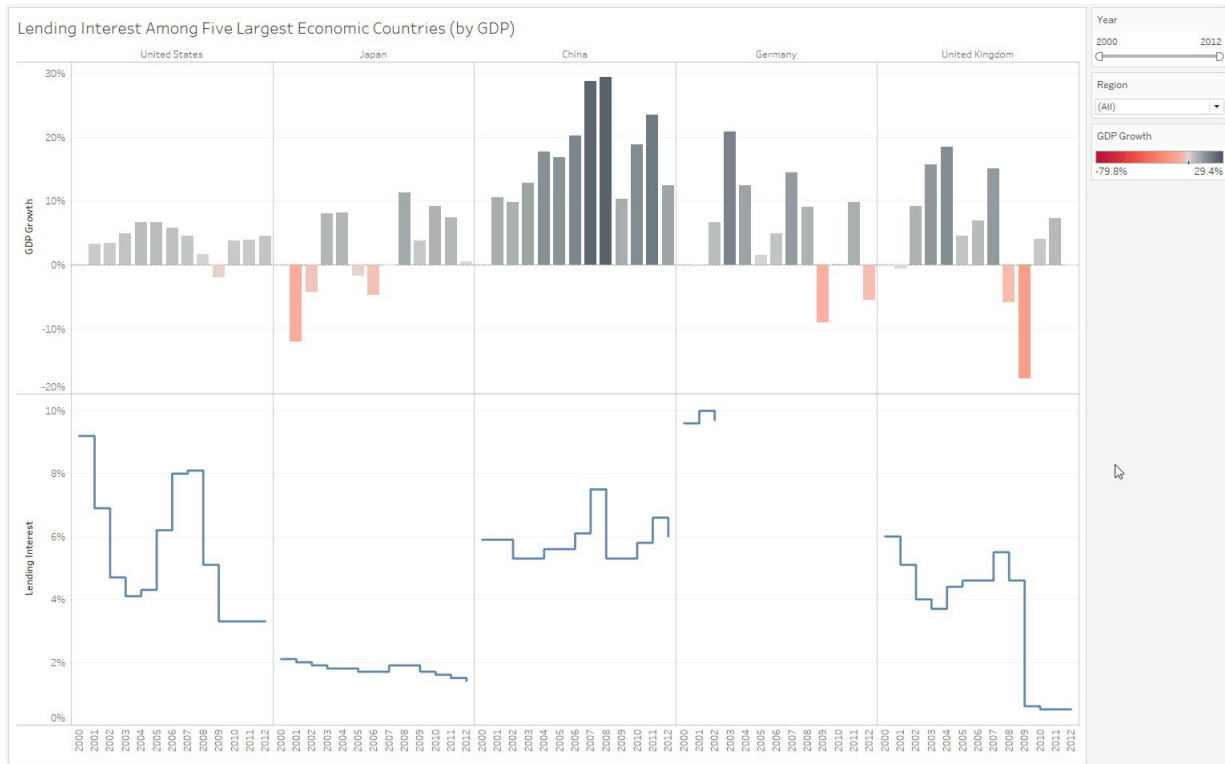


圖 8：使用 Tableau 的內建快速表格計算（百分比差異），即可快速呈現某些國家或地區中，利率提升對於 GDP 的影響。

4. 時間序列與預測性分析

功能：時間序列分析、預測

從感應器測量值、股市價格到畢業比例，世界上大部分的資料都能有效地建構成時間序列的形式。時間也因此成為分析專案中最常見的獨立變數之一。而如要獲得更多分析見解，您可能會想要運用預測相關功能。透過 Tableau，您不需要耗費大量心力，也能新增預測性分析，讓您快速探索多個情境。

時間序列分析

Tableau 的前端富有彈性，後端功能強大，您只要提出正確的問題，就能輕鬆地進行時間序列分析。只要將感興趣的欄位拖曳到視圖中，並啟動提問程序，就能開始進行分析。

您可以運用雙軸與離散化彙總功能，開始查看多個時間序列。如果無法以不同的細微度與彙總等級快速查看時間序列，您可能無法提出問題。

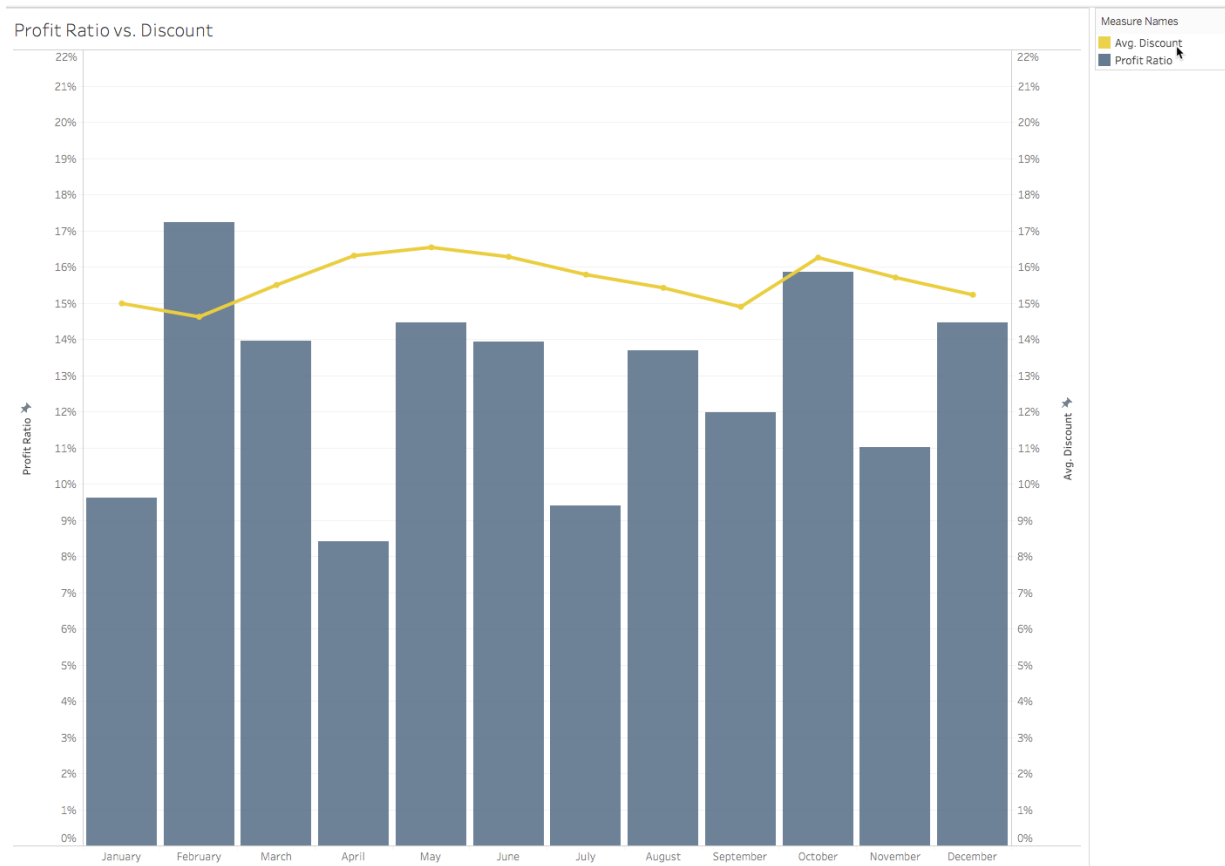


圖 9：此雙軸圖顯示利潤比例與平均折扣之間的關係。

如要查看某段特定的時間期間，可以使用一組確切的日期篩選資料，或是利用 Tableau 的相對日期篩選條件加以篩選。使用相對日期篩選條件時，可查看相對的時間期間，例如「上週」或「上月」。您每次開啟視圖來加入新資料時，這些時間期間都會更新，是非常適合用於報告用途的強大工具。

處理時間序列時，經常需要進行平滑處理，或執行其他時間相關計算。Tableau 提供一組豐富的功能，可簡化移動平均、與前一年比較的計算，以及累積總計等常見的時間序列運算。

如前文所述，Tableau 的表格計算功能提供一組常見的時間序列操控計算（快速表格計算）供您選擇，您也可以自行撰寫自訂運算。

Stock Tracking

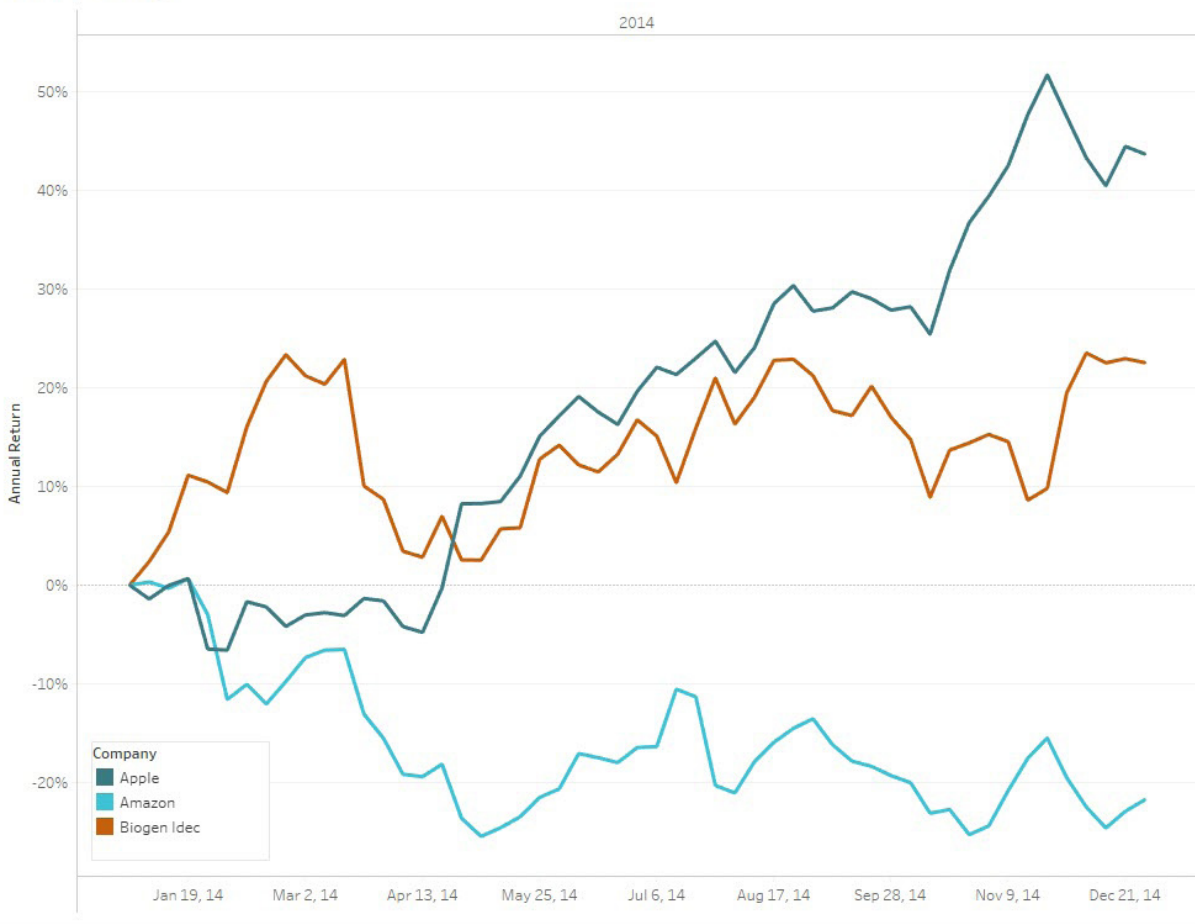


圖 10：此時間序列分析顯示三大公司的每月股票價格。

由於時間序列分析極為常見，Tableau 的功能可協助組織更快完成專案，並從專案中獲得更多價值。這個直觀易用的功能可幫助專家與商務分析師提出更多、更好的資料相關問題。

預測

如果需要預測未來情況，Tableau 備有數個原生建模功能，包含趨勢與預測的建模功能。

您可以在任何圖表快速新增趨勢線，而且只要以滑鼠右鍵點一下趨勢線，即可查看說明配適度的詳細資料（如 p 值與 R 平方）。因為趨勢線已完全整合入前端中，可輕易進行細分，因此只要使用 Tableau 的拖放功能，點一下滑鼠就能建立不同群組的模型。如圖 11 中所示，Tableau 無須任何程式碼，便會自動為各個不同的細分區段分別建立三條趨勢線。Tableau 也支援數種其他配適度功能，包含對數、多項式與指數。

Obesity's correlation with diet, exercise, and smoking (% of pop.) - All

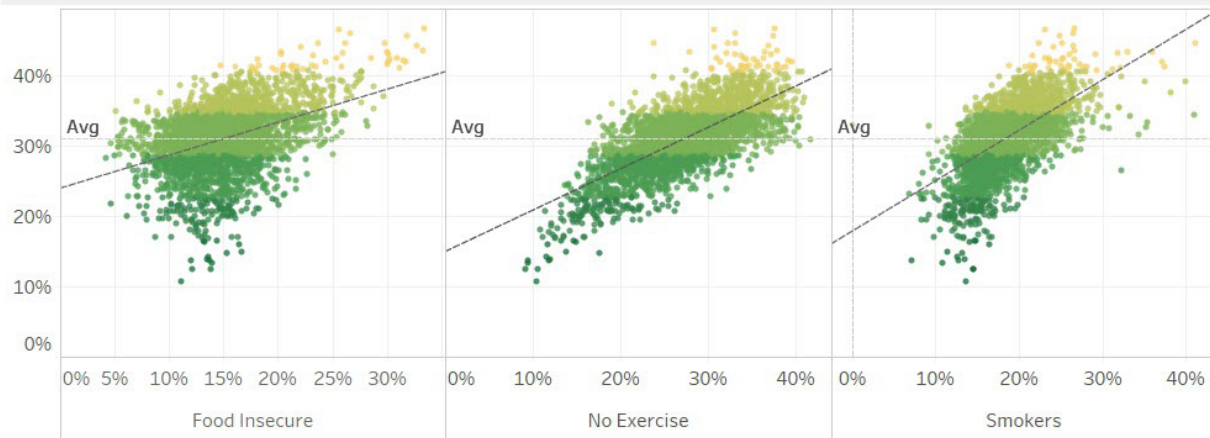


圖 11: 趨勢線標示出肥胖與食物取得不穩定、運動和抽菸人士之間的關係。

Tableau 的預測功能依預設會在背景執行數個不同的模型，然後選出最佳模型，而且會自動將季節性等資料問題列入考量。Tableau 中的預測功能使用一種稱為**指數平滑法**的技法。指數平滑法會使用過去資料值的加權平均，以迭代的方式預測時間序列的未來值。如前文所述，預測的各個層面，幾乎都能加以設定。舉例來說，您可以選取預測的時間長度、模型類型（相加或相乘），以及是否要將季節性列入考量。

這項功能非常容易使用，即使是新手使用者，只要使用預設設定並點幾下滑鼠，就能建立預測，進階使用者則能設定模型的幾乎所有層面。除了統計元素，Tableau 也會顯示信賴區間，為新手使用者提供預測品質的估計。預測功能也能與其他 Tableau 功能完美地搭配運作，因此預測亦能輕鬆地進行細分與操控，如同使用者介面中的任何其他分析物件。

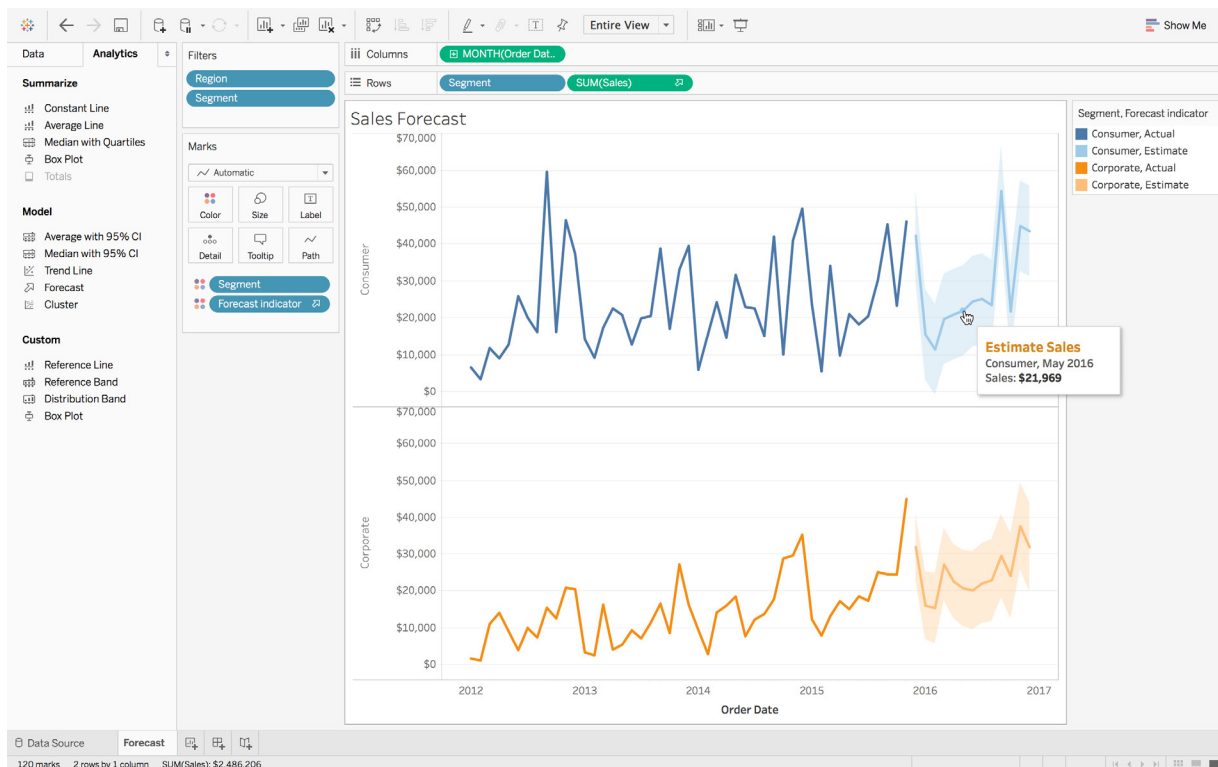


圖 12： 預測功能自動按細分區段預測銷售量。

能輕鬆進行預測性分析這項特點，可大幅增加幾乎任何資料專案的價值。Tableau 支援複雜設定與簡易互動式建模，是一個資料科學家還是最終使用者都能順利應用的平台。

5. 外部服務整合

功能：Python、R 與 MATLAB 整合功能

許多組織在分析平台與企業知識上進行投資，因此您可能會有一些非常具體的需求，以及一套具有價值的現有工作成品。

一個全方位的分析平台必須能夠支援整合其他進階分析技術，以便擴展可用的功能性，並運用其他解決方案中已投資的項目。運用 Tableau，您將能整合外部服務，如 Python、R 與 MATLAB 等，而且由於 Tableau 簡單易用，只要透過簡單的拖放操作，就能讓組織上下的所有使用者使用進階演算法，藉此實現資料科學大眾化。

Tableau 能直接與 Python、R 與 MATLAB 整合，透過現有模型與全球統計社群為使用者提供支援。Tableau 會將指令碼和資料傳送到 RServe 與 TabPy 等外部預測服務。結果接著會傳回到 Tableau，以供 Tableau 視覺化引擎使用。Tableau 使用者即可對於 Tableau 中的資料叫用 R 或 Python 中提供的任何函數，並使用 Tableau 操控在這些環境中建立的模型。

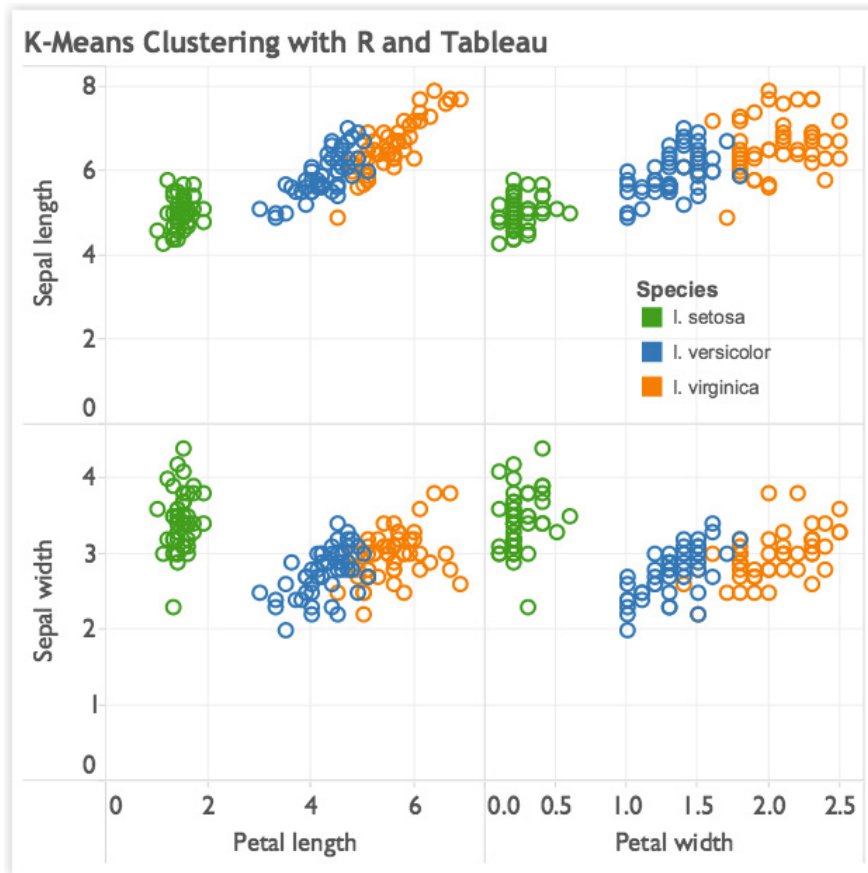


圖 13: 使用 R 的 k 平均聚類分析演算法範例，在 Tableau 中加以視覺化。
聚類顯示三種鳶尾花的變數（萼片與花瓣長度/寬度）差異。

建模作業能夠比基本統計還要深入許多。透過 R 與 Python 整合功能，您將能分析文字來擷取其中的情感（圖 14），或展示供應鏈或股票投資組合最佳化等規範型分析的結果。整合功能支援直接在 Tableau 中執行 R、MATLAB 或 Python 程式碼，也能直接在 Tableau 中執行於 Tableau 外定義的模型與指令碼。

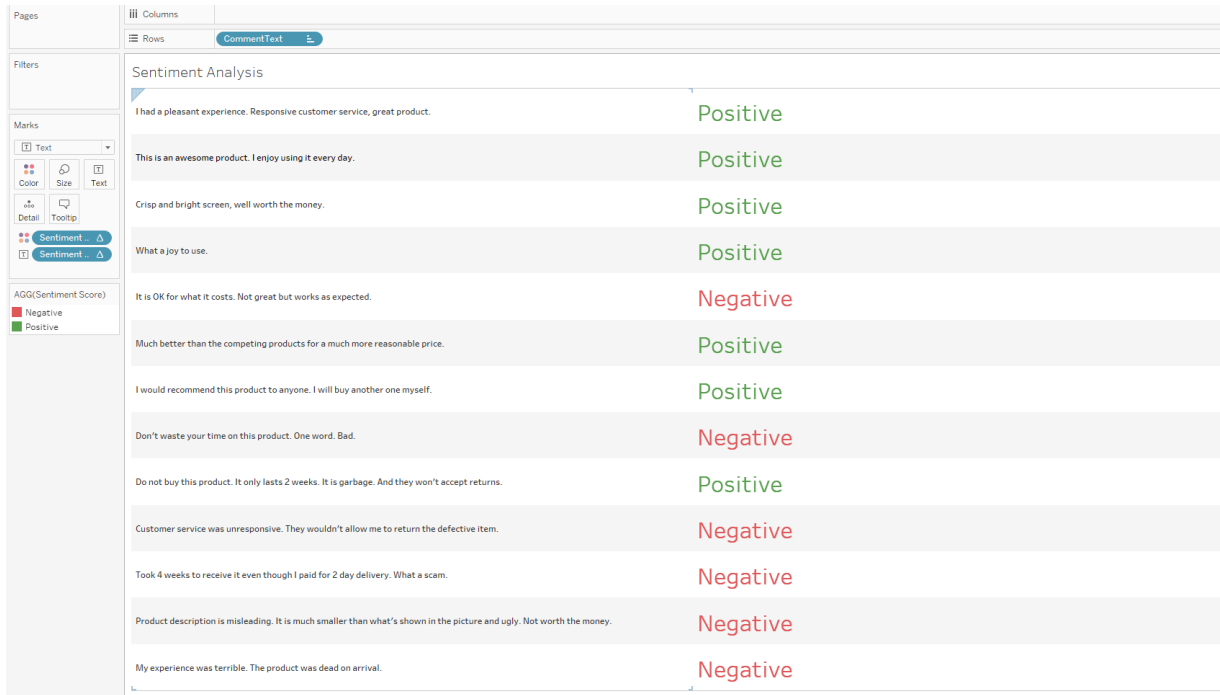


圖 14：使用 Tableau R 整合功能建立之情感分析的範例。

圖 15 顯示使用梯度提升的儀表板。這個儀表板使用梯度提升這項集成機器學習技法，根據過去的癌症案例提供胸部腫瘤的診斷建議。此模型的訓練與準確度統計評估，是由資料科學家在 Python 程式設計環境中進行，完成後再發佈到 TabPy (Tableau Python 伺服器)。這個儀表板提供表單輸入欄位，供使用者輸入不同醫療指標的值，這些值接著會再傳送到這個託管的機器學習模型，並即時轉換為診斷建議。這樣的作法可讓多個分析師和儀表板均使用共用的機器學習資產做為基礎進行建構，而資料科學家也能使用自選工具，在單一位置輕鬆管理這些機器學習資產。

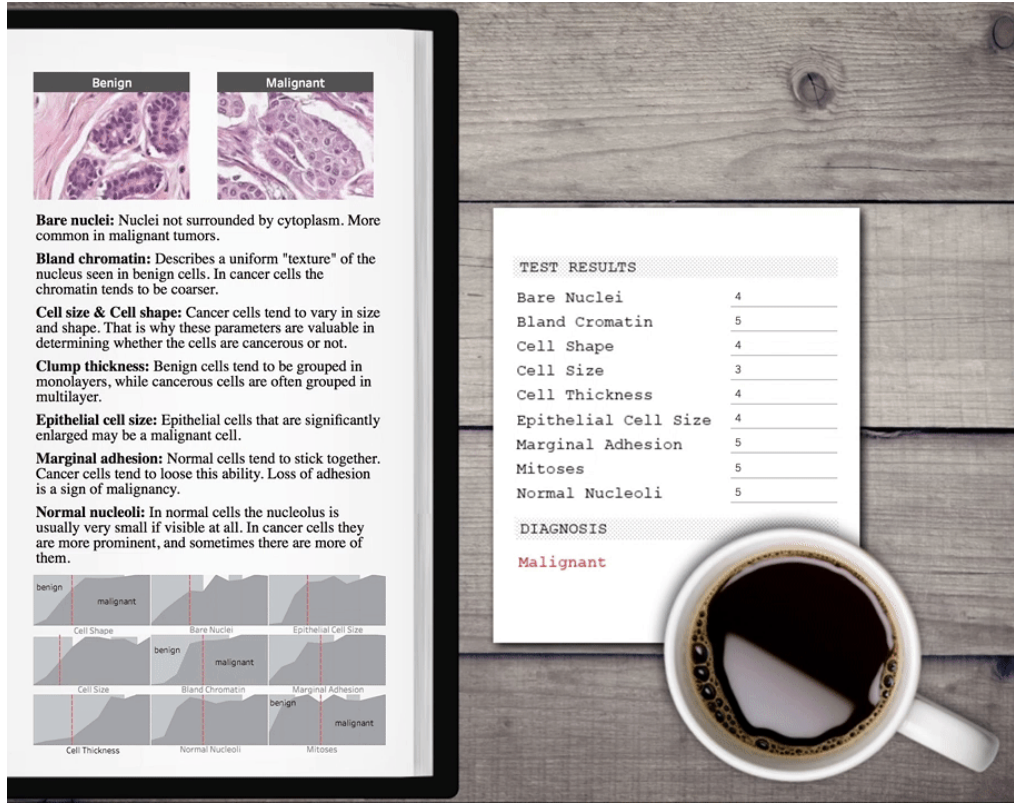


圖 15：具有輸入參數的集成機器學習模型。

將 Tableau 與 Python、R 或 MATLAB 結合運用，可發揮極為強大的功能。這樣的技術結合運用，能將複雜建模的結果呈現在精美且吸引人的視覺化呈現項目上，供任何技能水平的使用者使用。使用者將能透過儀表板上嵌入的控制項與資料互動，藉此提出假設性問題，並評估假設情境。

總結

在分析平台中，Tableau 有許多與眾不同的地方。由於擴增人類智慧是我們的使命，我們在設計 Tableau 時，已經將商務使用者與資料科學家納入考量。我們希望能讓使用者以最快速度，提出與資料相關的有趣問題，而我們打造的平台，正是我們持續致力實現這項使命的結晶，能針對各種技能水平的使用者提供有用的功能。

透過 Tableau 富有彈性的前端，商務使用者不必編寫程式碼或理解資料庫相關知識，即可提出疑問。Tableau 也具備資料科學家所需的分析深度，是資料科學家工具庫中的強大利器。藉由運用 Tableau 中的複雜計算、R 與 Python 整合功能、快速同類群組分析與預測相關功能，資料科學家可在 Tableau 完成複雜的分析，並輕鬆分享視覺化結果。無論是使用 Tableau 進行資料探索與品質控管，還是用於設計與測試模型，平台具互動性的性質，都能幫使用者在專案的生命週期期間省下無數的時間。Tableau 讓各種等級的分析都變得更平易近人，可更快速完成，進而促進組織內各處進行重要的協同作業，並做出更完善的決策。

關於 Tableau

Tableau 協助使用者從資訊中找出意義。Tableau 是一個支援分析週期的分析平台，提供視覺回饋，即使問題的複雜性持續演變發展，仍可協助您回答問題。如欲運用資料進行創新，應使用鼓勵您持續探索，持續提出新問題並改變觀點的應用程式。如果您已經準備好讓資料發揮影響力，歡迎立即下載 Tableau Desktop 的[免費試用版](#)。

相關白皮書

[使用 R 語言與 Tableau \(英文\)](#)

[瞭解詳細等級運算式 \(英文\)](#)

[定義分析 \(英文\)](#)

[查看所有白皮書](#)

瀏覽其他資源

[產品展示](#)

[訓練與教學課程](#)

[社群與支援](#)

[客戶案例](#)

[解決方案](#)

