

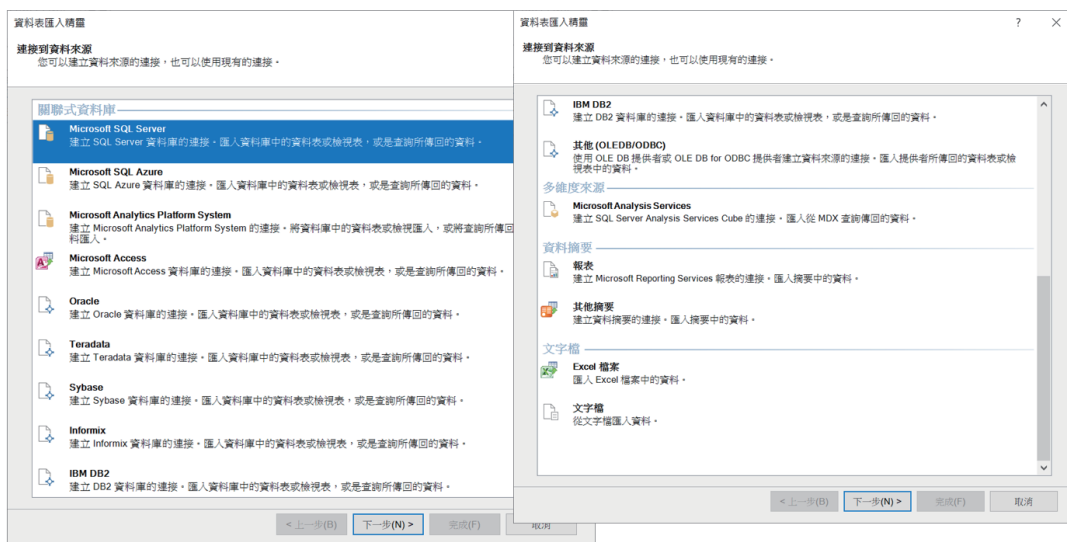
# 大數據分析 學習地圖

## 導覽

資訊或統計界的發燒話題，從大數據分析、物聯網到人工智慧（Artificial intelligence，簡稱 AI）等。許多學校跟企業也不斷地投入資源進行各種模組研發與探勘研究，無非是希望能找到新的商機應用。

然而筆者認為要邁向人工智慧的過程當中，最重要的就是大數據分析領域的蓬勃發展。全民大數據是一個相當重要的觀念，因為大數據分析技術有深有淺，分別可以應用在不同的目標階段上。為了要達到顧及全民大數據的目標，筆者從「關懷」角度切入來看，微軟的大數據系列是擁抱全民大數據和整合 Open Source 最廣且具有相當彈性，對於每一位想從事、想瞭解大數據相關的人來說是相當方便的。首先，最早微軟從 SQL Server 開始強化整合商業智慧服務（SSMS, SSIS, SSAS, SSRS）後，陸續在不同的服務或工具進行優化。

例如：Excel 2010 有 Power Pivot 後，到 Excel 2013 版之後加入 Power BI 增益集工具（內含 Power Pivot、Power Query、Power View 與 Power Map）；還有最近相當火紅的 Power BI（被 Gartner 評為 2019 年商業智慧與分析平台的領先者），以上兩者都屬於提供絕佳的資料視覺化分析；另外，從 SQL Server 2016 / 2017 開始，除了原有 T-SQL 外，也把資料科學家最常使用的 R 語言和 Python 語言整合進來，讓資料庫可以達到 end to end（資料庫→分析建模 / 預測分析 / 機器學習 / 深度學習→視覺化解讀呈現）的綜效。



↑ Excel Power BI 擁抱大數據

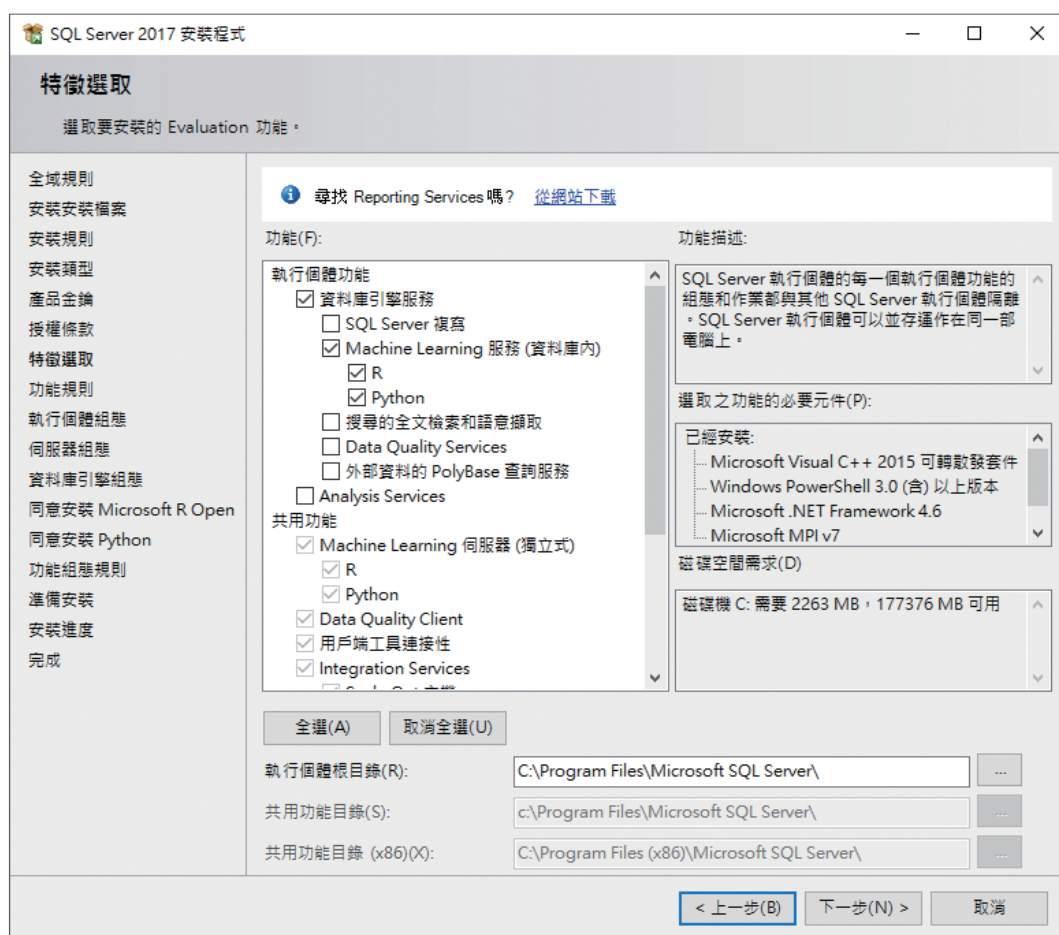


### ↑ Power BI 擁抱大數據

SQL、R 和 Python 已經變成現在資料科學家最重要的工具了，同時 Power BI 成為全民大數據在視覺化 BI 工具的首選。因此，對於以實務商業行銷管理應用為主的資料科學家們來說，同時具備這 3 種語言能力和操作 Power BI 工具是非常重要地。

若要區分 R 和 Python 不同差異，筆者整理如下：

- 分析建模／預測分析／機器學習／深度學習 可選擇 R 語言。提供優異靈活性與思考，內建齊全的統計和數據分析圖形，還有持續更新及增強分析的功能，讓它可以擴增到許多領域來應用。
- 數據預處理清洗（非結構化數據，例如網頁、圖片、文字等）／機器學習／深度學習可選擇 Python 語言。這方面提供強大的靈活性，從社群／網站／媒體爬取訊息後的分析彈性；處理大量的數據能力會比 R 語言好。

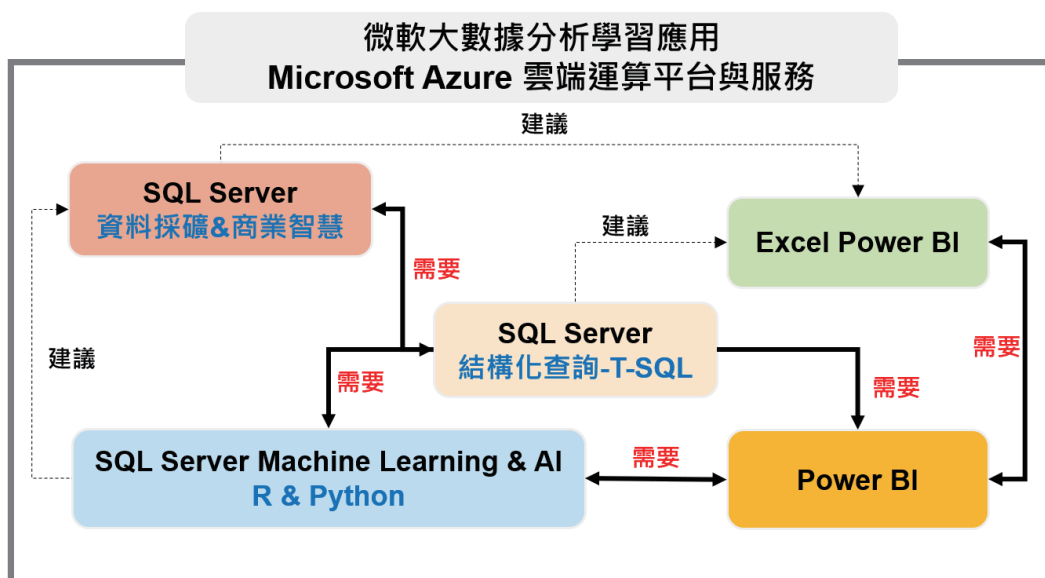


↑ SQL Server 2017 – Machine Learning With R & Python

不過對全方位的資料科學家來說，應該都要有程度上的瞭解 SQL / R / Python / 視覺化 BI 工具各自的優點，因為在面對分析的各項任務過程當中，我們需要的是這些語言工具（SQL / R / Python / Power BI）的幫忙，應用在不同任務上來解決問題，並非去比較它們。

## 大數據分析學習系列

微軟的大數據分析服務可說是相當完整，筆者為此擬出一學習地圖，從分析角度和全民適用等 2 個方向來學習大數據分析。讓尚未進入大數據或剛進入大數據的資料分析人員能有一個學習參考。



▲ 微軟大數據分析學習地圖

# Power View 資料探索 智慧視覺

## 3-1 資料視覺化：Power View

---

什麼是 Power View？它是一種整合網頁技術 Silverlight 的視覺化工具。Power View 出現是微軟商業智慧架構中，最能夠讓人感受到什麼是自助式 BI，更替 Power BI 元素增添畫龍點睛效果。

早期微軟商業智慧的 Reporting Services 有報表產生器，可是它對於完全沒有 IT 背景的人員來說，使用起來依舊存在一定的門檻。不過自從 Power View 出現後，對於報表工具的定義使用卻有很大突破，重點是夠直覺、簡單且易操作，使用者可透過互動式圖表、圖形功能來進行資料探索（explore）工作，並將結果利用視覺化方式來呈現。這與市面上目前許多商業智慧工具都已強調視覺化、資料視覺化或視覺化分析等，甚至巨量資料視覺化的觀念，其實都是在闡述同一件事情。

簡單說，使用 Power View 最終目的就是希望能快速地將資料透過圖表來呈現，清楚地傳達資料趨勢和分佈結構，操作過程著重簡單直覺，當然這些視覺化 BI 工具的主要使用者（或稱取用者），大多是來自企業高階幕僚或行銷業務人員；對他們而言，快速產生決策是他們最在意的事情，因此呈現結果和快速理解，就是視覺化 BI 工具的任務了！

Power View 的視覺效果取決於 Silverlight 技術核心，所以使用 Power View 除了需安裝 Excel 2013 之後版本（Office 365 亦提供 Power BI 功能）之外，同時要有 Silverlight 5（下載網址：<http://www.microsoft.com/silverlight/>）。

立即存取、直覺操作、建立視覺化報表和呈現，這些完全都可透過 Excel 來解決，且對於公司內部不同職掌的人或角色來說，亦可共享 Power BI 功能。對於使用者角色而言，

通常區分為報表建立者／使用者、資訊管理者或資訊專業人員以及上述提到的報表取用者—中高階幕僚，大致扮演的角色功能如下：

### ● 報表建立者／使用者

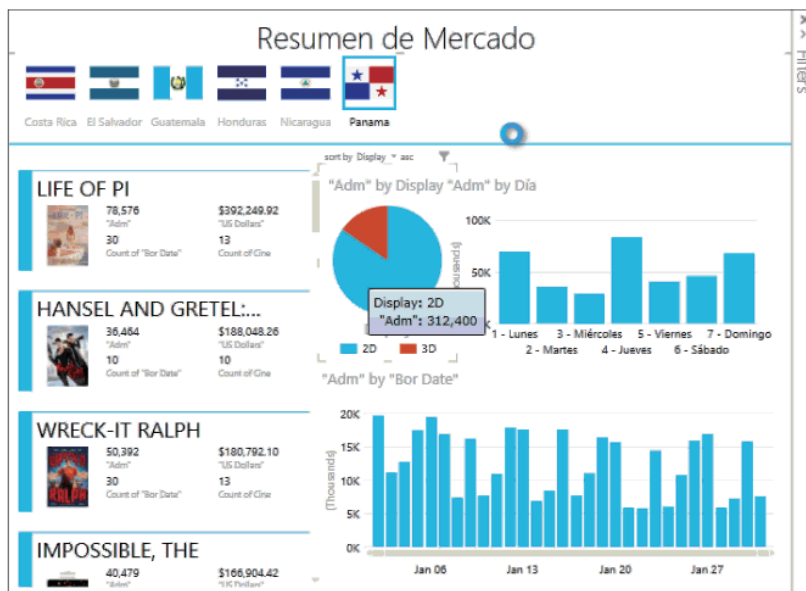
通常是指數據分析師、資料採礦分析師、BI 使用者或一般使用者，因為工作上需要的關係，這些人必須不斷進行資料探索、建立企業管理報表或業績報表等。也由於他們常需要透過探索資料的關係，故必須親自建立報表，理所當然是報表的使用者之一。

### ● 資料管理員或資訊專業人員

須對 Power BI 相關服務功能非常熟悉瞭解，因為一般 BI 工具會有特定與安全管理的資料源存取、權限管理核准項目、資料部署及安全發佈等問題，故對企業來說相當重要。這群人必須熟稔 Power BI 相關佈建資訊。

### ● 報表取用者

廣義而言，任何人都可以是報表取用者，不過以企業組織定位來看，通常指的是幕僚決策者。原因是這些人有決策執行權，並且有一定程度的商業分析能力。



▲ 圖 3-1 Power View 視覺化範例圖（參考來源：<http://www.poweredsolutions.co/>）

### 3-3-4 進階視覺功能應用

#### 3-3-4.1 下鑽（Drill down）

相信讀者對於下鑽（Drill down）這個名詞應該不陌生。只要是屬於維度階層的關係欄位，就可進行下鑽（Drill down）操作。嚴格來說，在執行下鑽（Drill down）之前的定義必須是直覺且具分析意義，也就是階層配置須經過周詳思考。

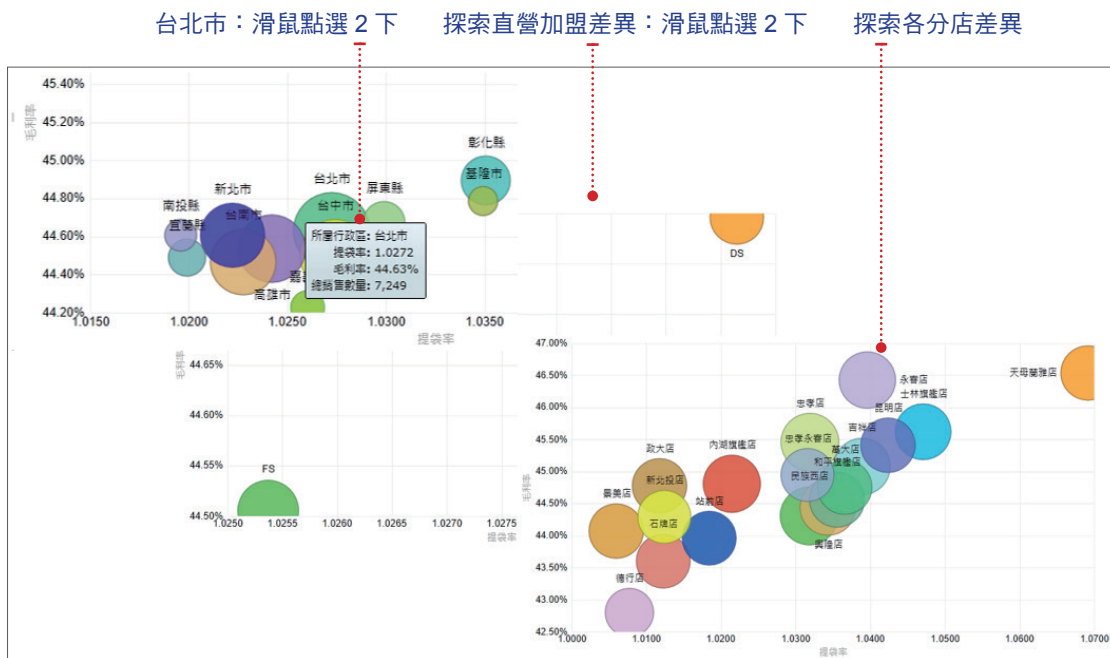
在 Power View 中，同樣具有與市面上商業智慧工具一樣的下鑽功能，主要區分為兩種方式，依序為已設計階層屬性和未設計階層屬性。

#### ■ 已設計階層屬性



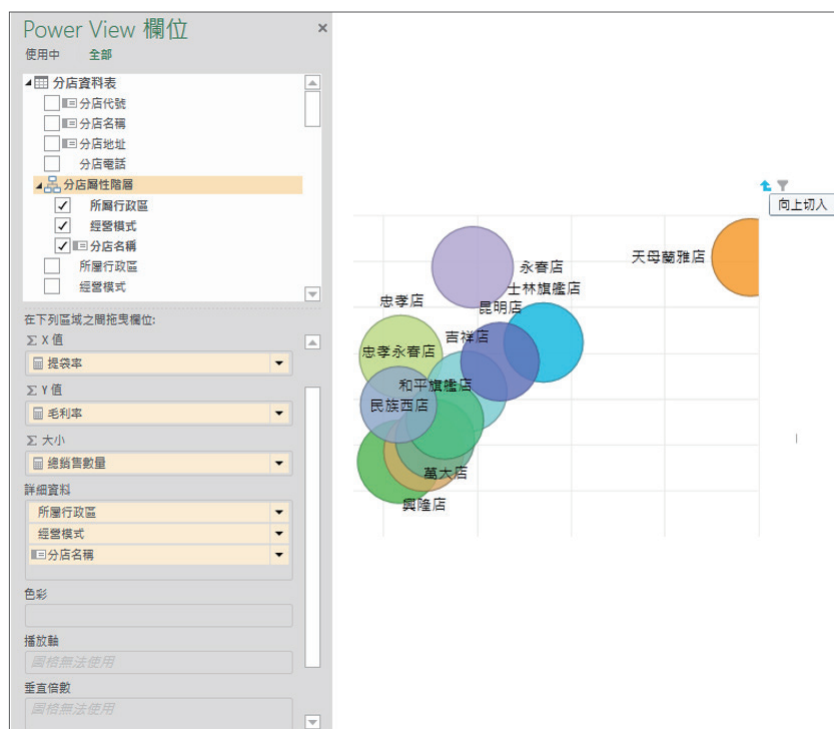
使用方式：只要欄位配置區選擇該階層屬性欄位，內含的任何資料行組合都可套用下鑽（Drill down）功能

- 1 沿用上述散佈圖案例進行介紹。因欄位配置區已存在分店屬性階層，所以直接點選球形圖兩下進行下鑽動作。



- 2 若在資料探索過程當中，欲回到上一層階層屬性時，請點選圖表右上角向上切入（roll up）按鈕即可。基本上，非資料表格類型物件（指圖形物件）的資料探索

過程中，其執行向下鑽研（drill down）和向上切入（roll-up）的操作步驟非常相似，讀者可自行多加練習。



## ■ 未設計階層屬性



使用方式：可任意選擇資料行組合套用下鑽（Drill down）功能

- 1 設定標題為各縣市產品類別銷量之毛利率和提袋率（同時請自行調整格式）。
- 2 為能夠讓讀者清楚瞭解未設計階層屬性時，如何執行向下鑽研（drill down）和向上切入（roll-up）動作，筆者使用資料表格類型物件（指非圖形物件）來說明。於欄位配置區依序勾選或拖曳需求欄位：所屬行政區、分類名稱、品牌名稱、年齡級距與總銷售數量。

在下列區域之間拖曳欄位：

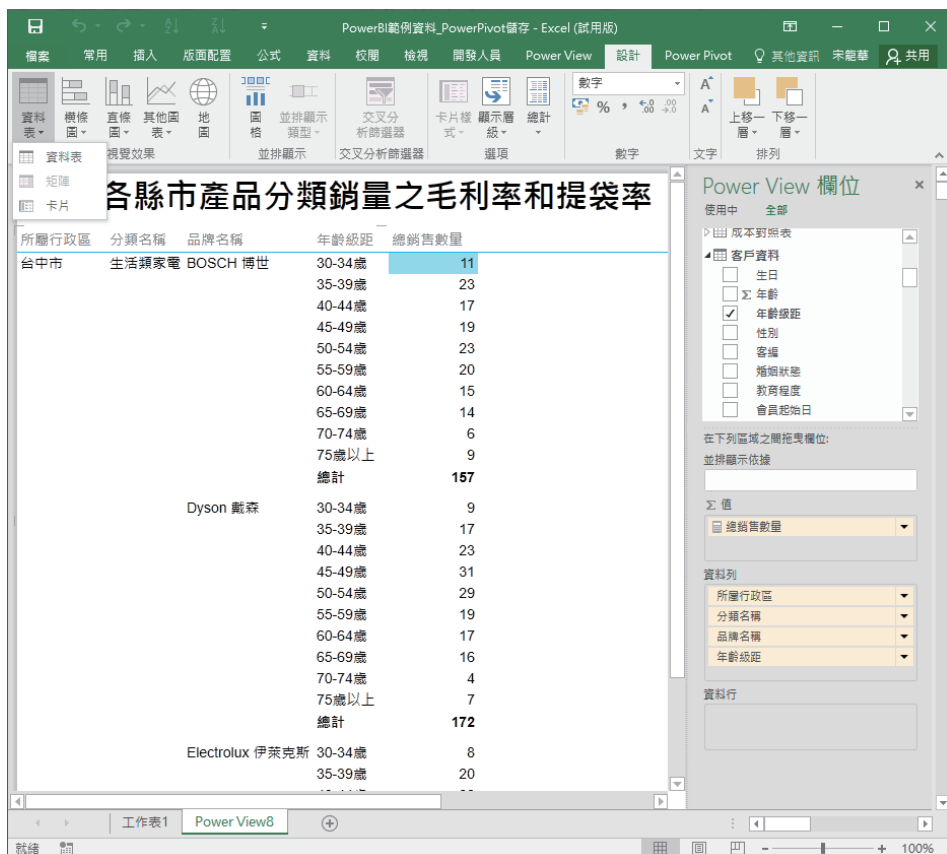
並排顯示依據

欄位

- 所屬行政區
- 分類名稱
- 品牌名稱
- 年齡級距
- 總銷售數量



### 3 切換成矩陣型式。



### 4 於設計索引標籤下選擇顯示層級資料列－啟用向下鑽研（一次一層）。



### 5 將滑鼠移至所屬行政區，畫面會出現向下鑽研指示，如此就可進行資料探索動作。

- 6 點選向下鑽研後，在同個畫面區域內，一樣會出現向上切入，同樣可回到上層屬性層級。

所屬行政區	總銷售數量
台中市	14,922
台北市	17,070
台東縣	820
台南市	12,650
宜蘭縣	453
宜蘭縣	3,170
南投縣	2,088
屏東縣	4,477
高雄市	12,958
基隆市	1,585
新北市	11,854
嘉義市	2,057
嘉義縣	2,472
彰化縣	6,491
總計	93,067

分類名稱	總銷售數量
生活類家電	6,264
個人類家電	2,967
健康美容類家電	1,391
廚房用小家電	4,300
總計	14,922

- 7 將年齡級距欄位移至資料行。

在下列區域之間拖曳欄位:

並排顯示依據

Σ 值

總銷售數量

資料列

所屬行政區

分類名稱

品牌名稱

資料行

年齡級距

- 8 形成類似交叉表的效果。我們可以同時進行資料探索，並一目了然知道各縣市區域下的產品分類或品牌名稱，在各年齡級距的分佈情況。

比較各縣市產品分類銷量之毛利率和提袋率

所屬行政區

30-34歲

35-39歲

40-44歲

45-49歲

50-54歲

55-59歲

60-64歲

65-69歲

70-74歲

75歲以上

總計

台中市

760

1,504

2,253

2,333

2,224

1,894

1,383

1,046

687

838

14,922

台北市

1,729

2,508

2,649

2,481

2,162

1,603

1,169

784

977

17,070

台東縣

46

77

132

125

131

106

75

55

29

44

820

台南市

674

1,312

1,900

1,991

1,882

1,558

1,144

884

590

715

12,650

宜蘭縣

35

34

66

77

62

49

45

30

27

28

453

花蓮縣

194

299

479

487

486

413

291

226

123

172

3,170

南投縣

119

210

294

325

312

240

203

137

98

150

2,088

屏東縣

246

453

650

706

639

571

415

302

249

246

4,477

高雄市

685

1,362

1,934

1,987

1,973

1,611

1,173

887

597

749

12,958

基隆市

84

148

230

285

226

192

122

110

68

120

1,585

新北市

674

1,290

1,703

1,709

1,719

1,517

1,051

815

520

640

11,954

嘉義市

嘉義縣

彰化縣

分類名稱

30-34歲

35-39歲

40-44歲

45-49歲

50-54歲

55-59歲

60-64歲

65-69歲

70-74歲

75歲以上

總計

生活類家電

314

626

963

980

940

828

580

422

274

337

6,264

季節類家電

319

453

462

433

363

252

210

145

177

2,967

健康美容類家電

廚房用小家電

品牌名稱

30-34歲

35-39歲

40-44歲

45-49歲

50-54歲

55-59歲

60-64歲

65-69歲

70-74歲

75歲以上

總計

BOSCH 博世

11

23

17

19

23

20

15

14

6

9

157

戴森

9

17

23

31

29

19

17

16

4

7

172

Electrolux 伊萊克斯

8

20

22

22

21

17

11

12

6

2

141

EMEME 鴻奇

5

13

23

35

33

19

13

15

7

9

172

GREE 格力

9

6

32

20

18

26

13

11

5

4

144

HERAN 禾聯

7

21

15

17

13

25

12

19

8

6

143

HITACHI 日立

36

68

92

92

99

84

56

33

30

35

625

HITACHI 日立

7

20

26

20

17

23

14

9

5

8

149

IZUMI 捷尼思

11

21

31

21

21

25

20

9

11

13

183

總計

22

47

22

23

21

20

20

20

15

17

1,172

### 3-3-4.2 交叉分析篩選器

除了 Power Pivot 之外，Power View 也有交叉分析篩選器功能。我們在上述內容中已有初步說明利用並排顯示選項的索引標籤帶和圖格流程功能來進行篩選，可是這些篩選功能基本上仍存在某些限制：僅能在該圖表物件或部分圖表物件之下進行篩選，不能針對整個 Power View 報表內的所有圖表物件進行篩選。

因此篩選效果若要擴大到整個 Power View 報表時，須使用內建功能交叉分析篩選器（Slicer）。它的功用與上個章節 Power Pivot 的交叉分析篩選器概念其實一樣，皆是針對資料過濾篩選後，呈現顯示自己想要的資料，以利執行更多的維度資料分析。Power View 的交叉分析篩選器（Slicer）屬於篩選控制項，可縮放顯示在 Power View 報表內的範圍，並進行互動式顯示資料變更。



使用方式：設計索引標籤下選擇交叉分析篩選器（Slicer），將資料行轉換成篩選控制項



↑ 交叉分析篩選器設定處

- 1 沿用上述散佈圖案例進行介紹。移除原有並排顯示依據的分類名稱欄位，這時原有的圖格篩選功能也會一併消失

在下列區域之間拖曳欄位：

並排顯示依據

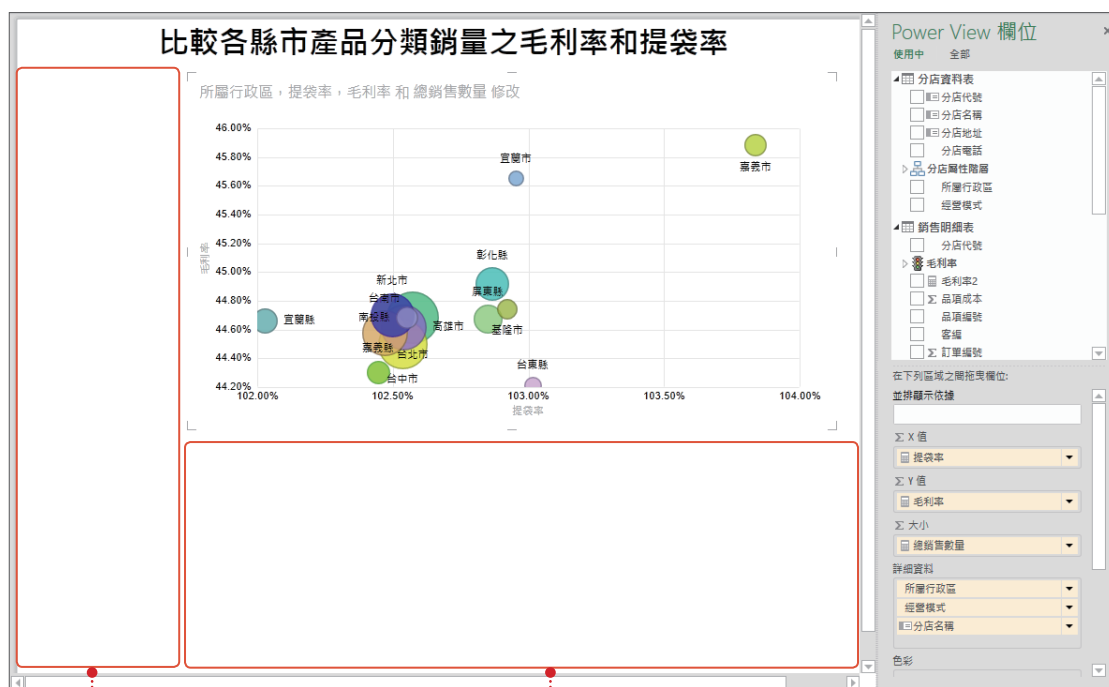
Σ X 值

Σ Y 值

Σ 大小

詳細資料

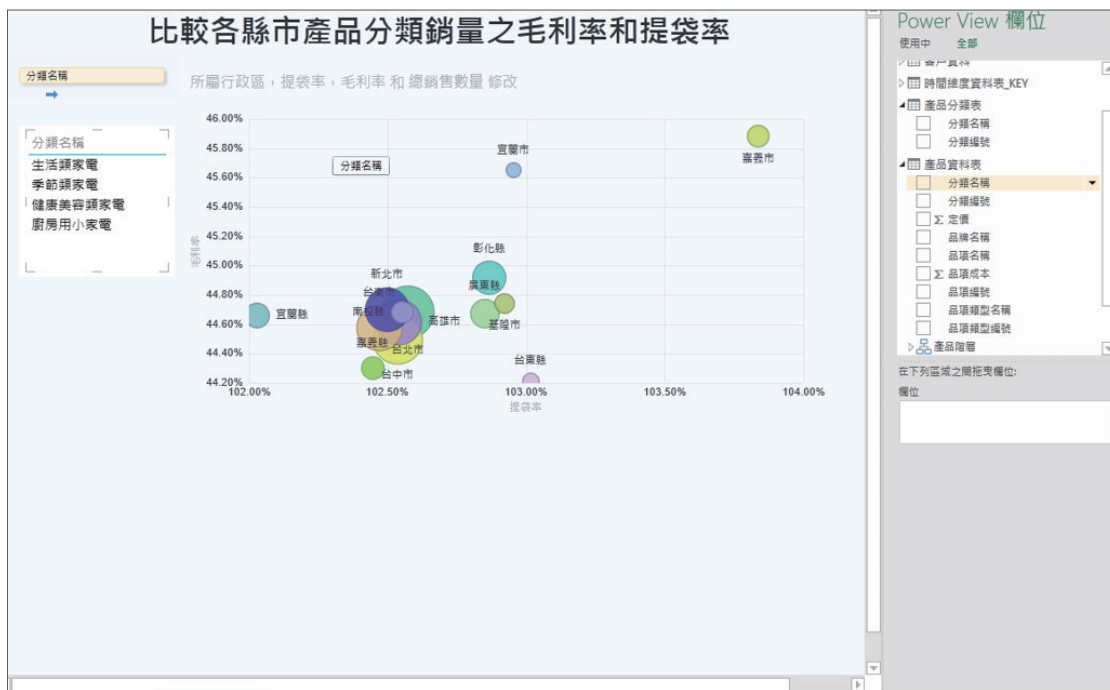
- 2 將版面調整至偏右狀態，以利後續擺放篩選器控制項。



預留擺設 [ 交叉分析篩選器 ]

可新增 - 圖表物件空白區

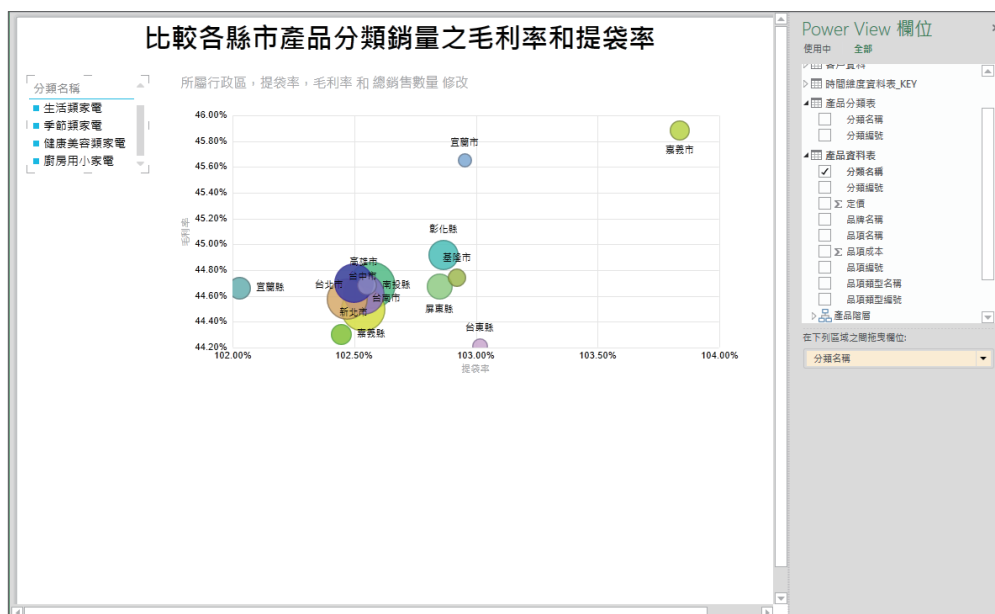
- 3 利用滑鼠拖曳方式將分類名稱欄位拖曳至預留擺放處，此時會自動形成篩選器控制項物件雛型。



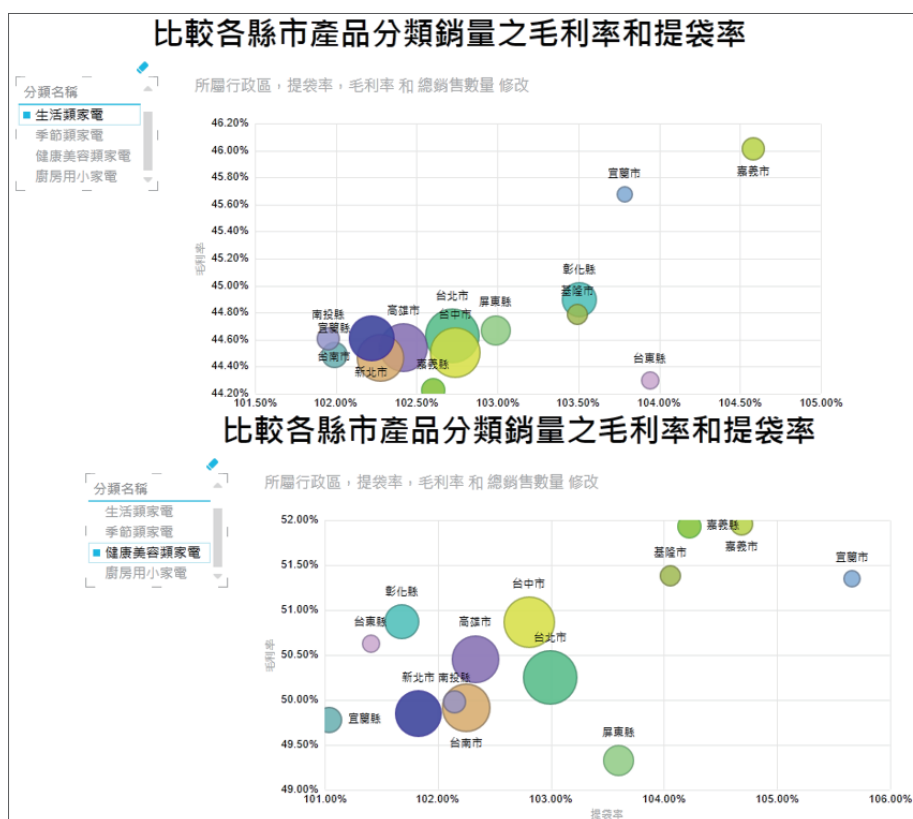
- 4 選擇該篩選器控制項物件雛型後，點選設計索引標籤下的交叉分析篩選器（Slicer），正式形成交叉分析篩選器控制項物件。



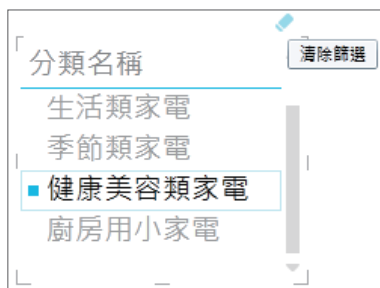
5 完成後請自行調整控制項顯示外觀。



6 可自行選擇控制項內的其中一項分類名稱，進行互動式的顯示資料變更。



- 7 倘若要清除篩選器控制項物件內的篩選條件，則點選篩選器控制項物件右上角圖示，即可清除並回到整體資料結果。



讀者可自行再多做變化。例如新增品牌類型名稱維度和另一圖表物件，同樣置換成交叉分析篩選器，依舊可進行互動式的顯示資料變更，等於是再增加探索資料的維度面向。如下圖是增加探索資料的維度（例如品項類型名稱）。

