期末報告 雲端運算與分散式資料庫

OLAP線上分析處理

嚴重特殊傳染性肺炎分析

指導教授：蔡正發

系級：資訊管理系三A

學生：B10756038施宗佑

報告日期：2021年6月11日

目錄

[一、研究動機](#_Toc72337433) 3

[二、資料集介紹](#_Toc72337434) 3

[三、資料預處理](#_Toc72337435) 3

[四、切片(Slice)分析](#_Toc72337436) 11

[五、切丁(Dice)分析](#_Toc72337437) 12

[六、向下擷取(Drill-Down)與向上截取(Roll-Up)分析](#_Toc72337438) 19

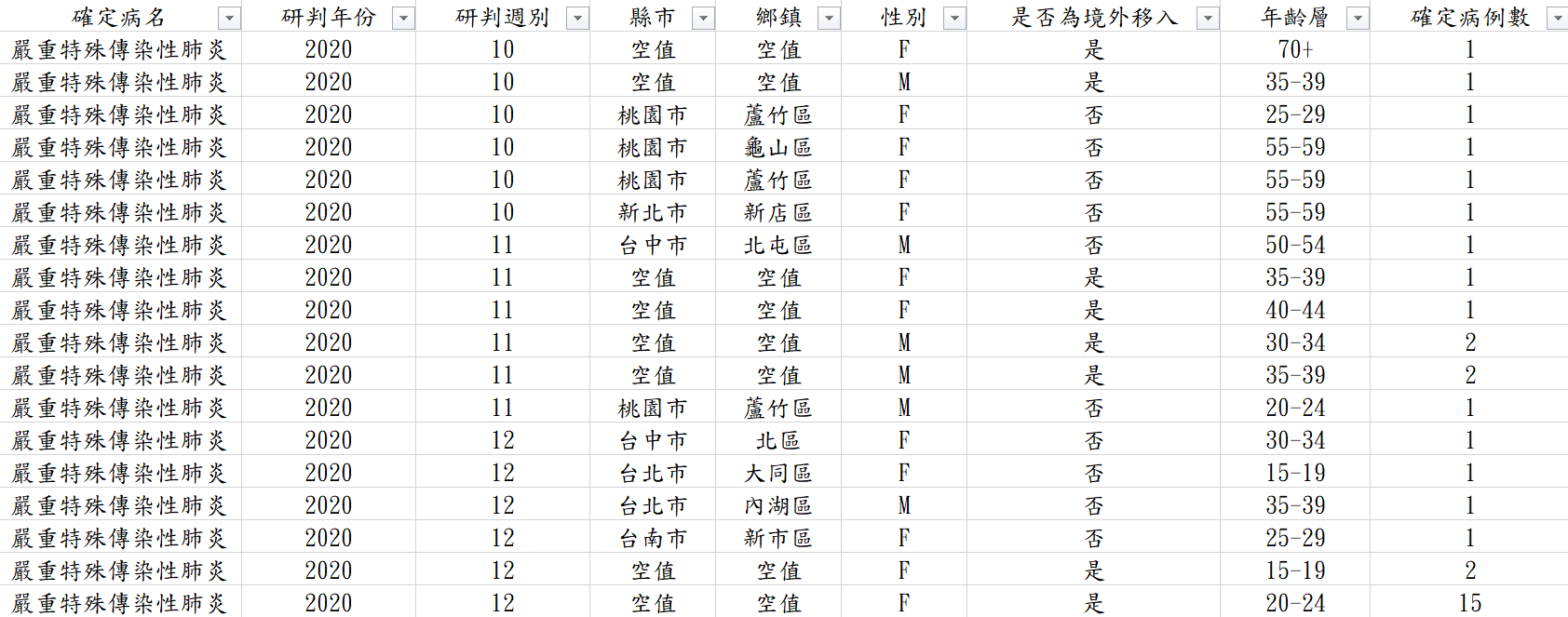
[七、轉軸(Pivot)分析](#_Toc72337438) 20

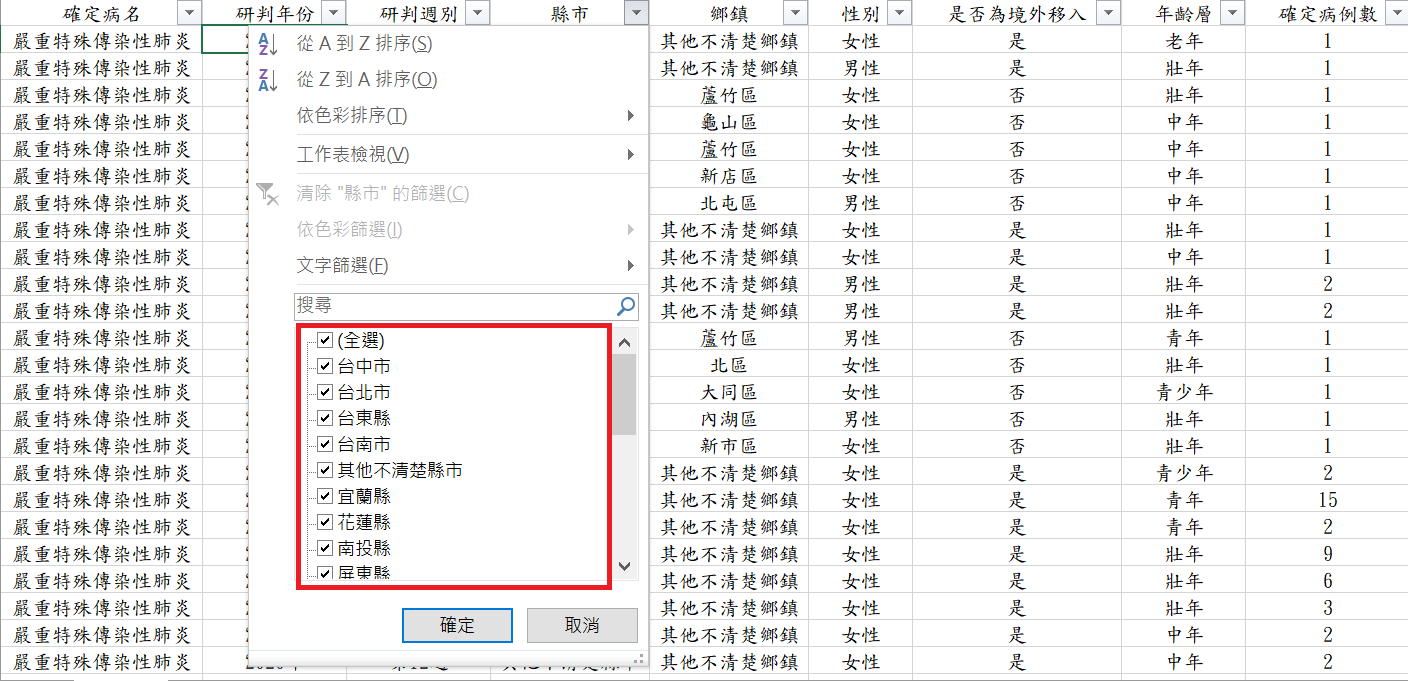
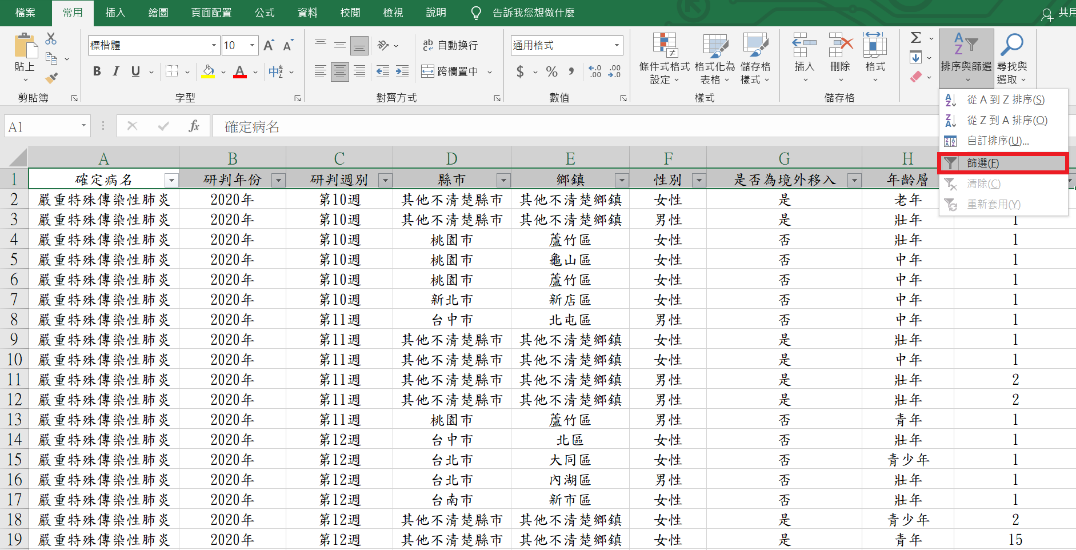
[八、分析結果](#_Toc72337438) 20

[八、心得](#_Toc72337438) 21

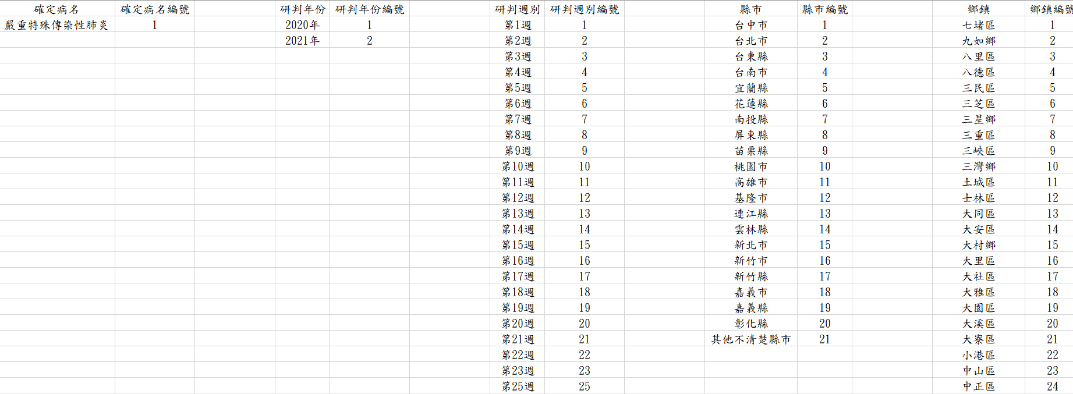
1. 研究動機

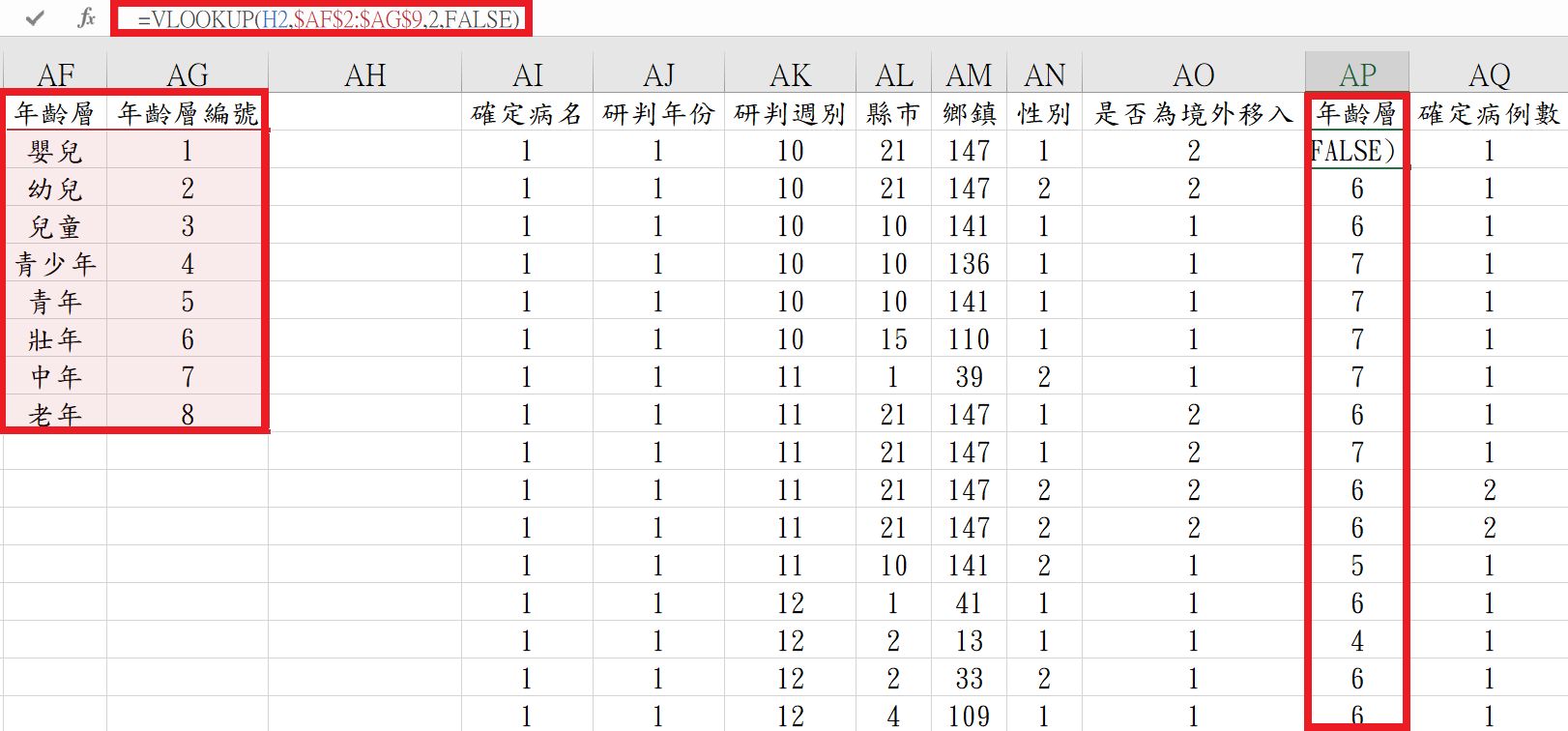
隨著2021年5月中爆發新一波社群感染的新冠肺炎人數，透過OLAP分析出其中的地區關聯性，為了製作此份報告，將資料更新到2021年5月27日，此週期為肺炎疫情人數的高峰，以提高精確度。

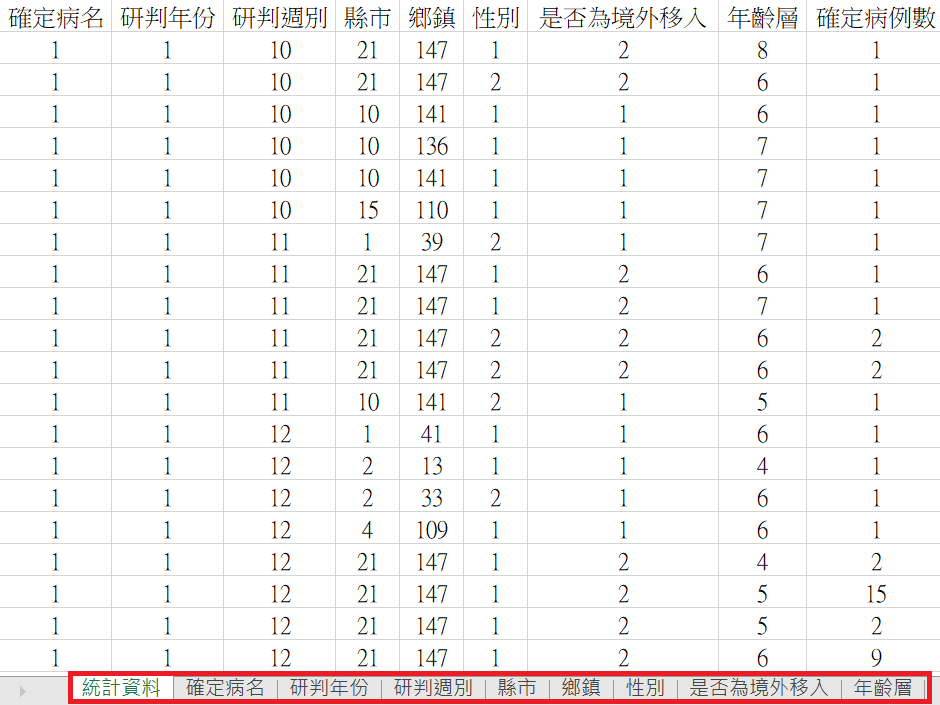
1. 資料集介紹
   1. 資料集名稱：地區年齡性別統計表-嚴重特殊傳染性肺炎(以週為單位)-依個案研判日統計
   2. 資料來源：疾病管制署資料開放平台
   3. 資料連結：<https://data.cdc.gov.tw/dataset/aagsdctable-weekly-19cov>
   4. 資料背景-各Table資料量
2. 確定病名：1筆(嚴重特殊傳染性肺炎)
3. 研判年份：2筆(2020年~2021年)
4. 研判週別：52筆(第1週~第23週、第25週~第53週)
5. 縣市：21筆(全台縣市與其他不清楚縣市)
6. 鄉鎮：147筆(全台鄉鎮與其他不清楚鄉鎮)
7. 性別：2筆(女性與男性)
8. 是否為境外移入：2筆(否與是)
9. 年齡層：8筆(從嬰兒到老年區分8個層級)
10. 確定病例數：6088筆(確診人數)
    1. **此份即時資料更新至2021年5月27日，以利報告的精確度。**
11. 資料預處理
12. 原始資料
13. 資料處理過程

**步驟一**：利用Excel中**排序與篩選**裡的**篩選**將欄位內的資料篩選出來。

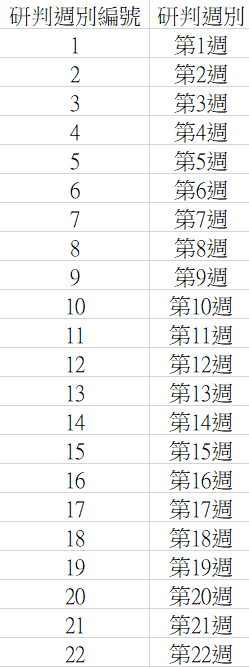
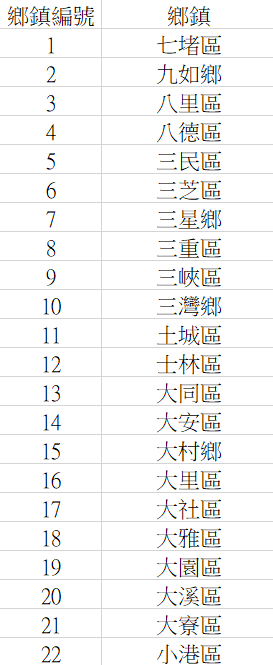
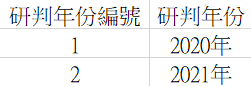
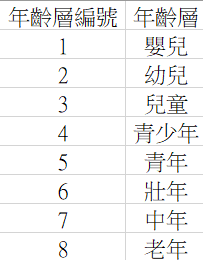
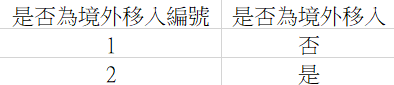
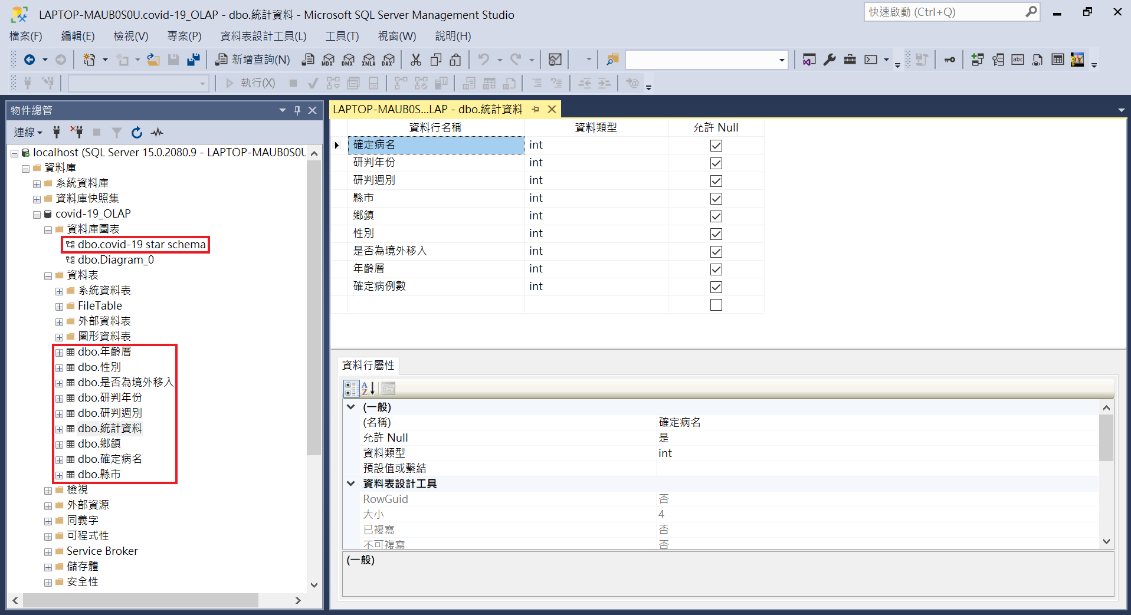
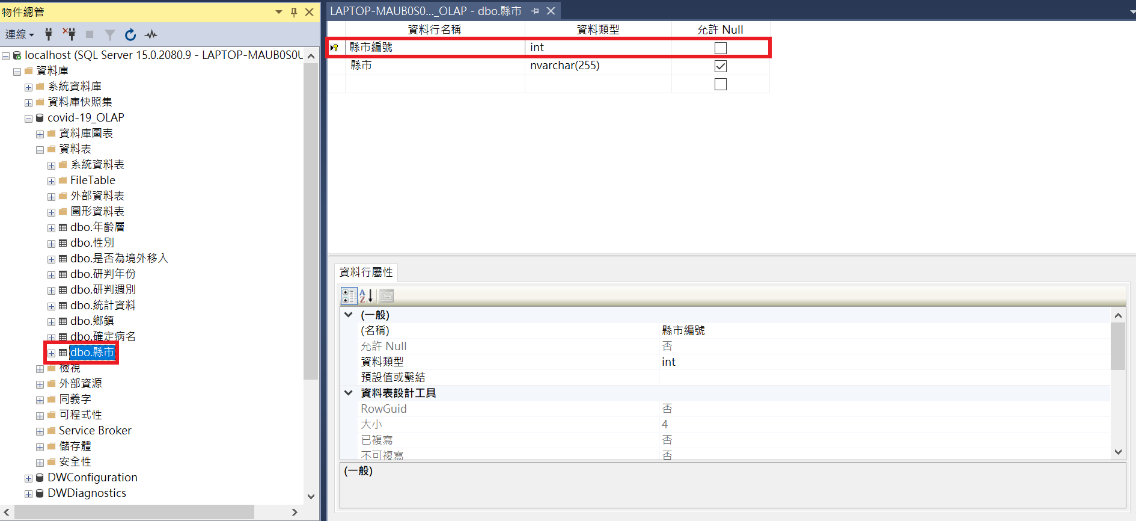
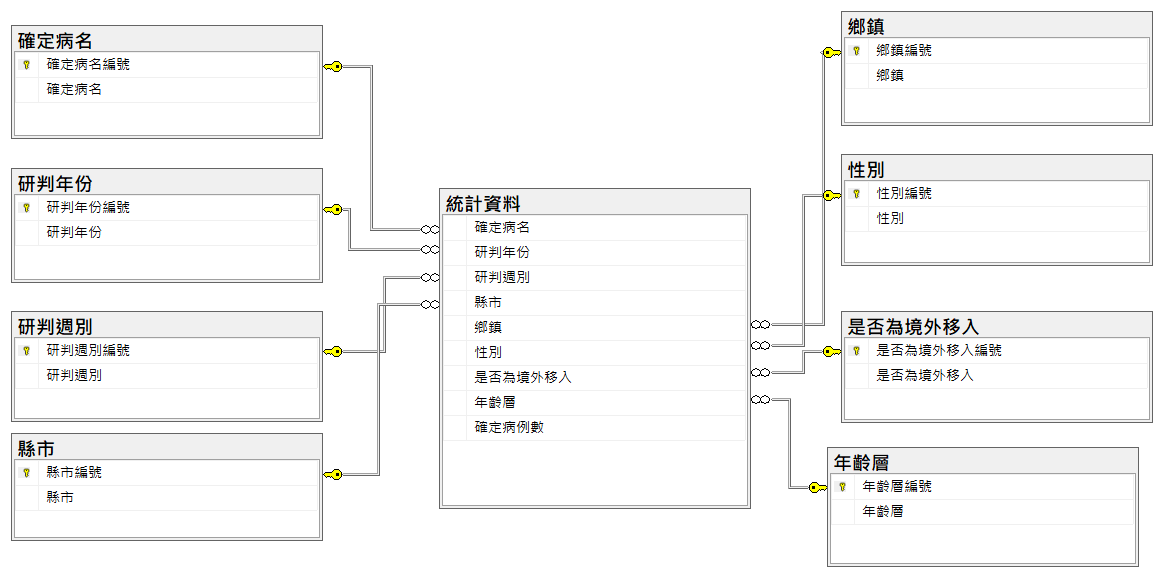
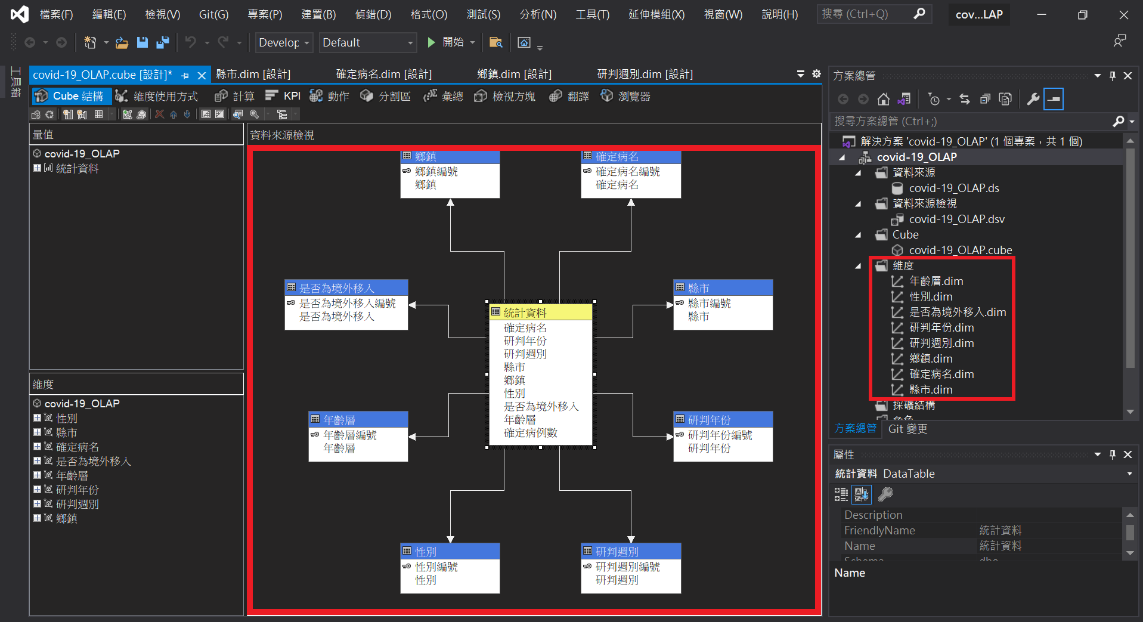
**步驟二**：欄位內篩選出的資料。

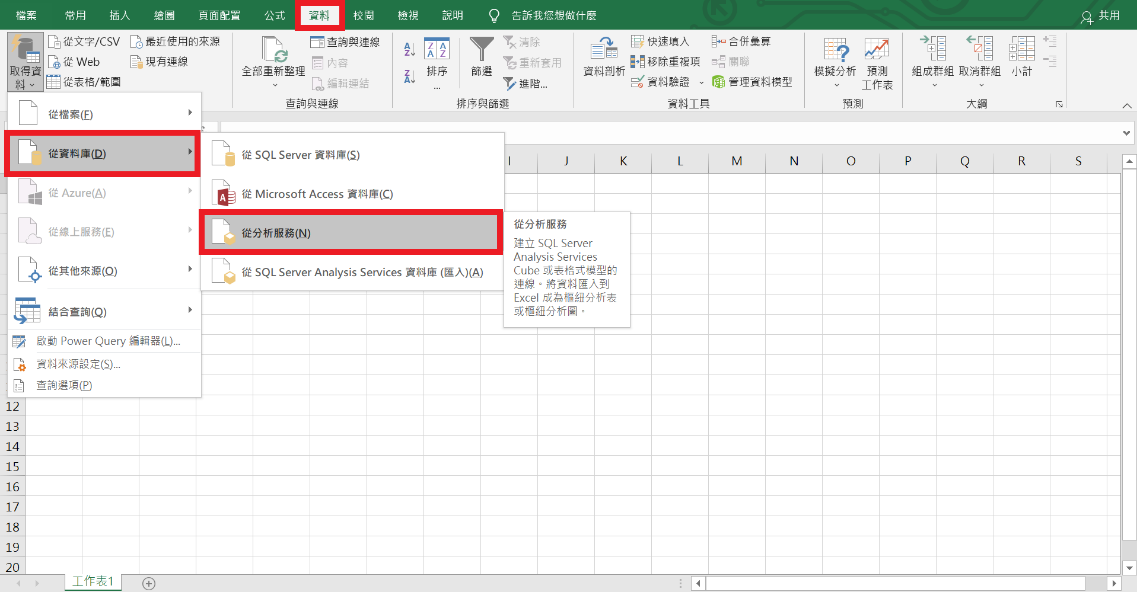
**步驟三**：將欄位內篩選出的資料依序給予編號，以此類推完成欄位轉換。

**步驟四**：利用Excel中**VLOOKUP公式**轉換成設定好的排序資料，以**年齡層欄位**來說，如果欄位H2(老年)利用VLOOKUP公式鎖定分析欄位AF2至欄位AG9，符合AF9(老年)給予AG9(8)值，以此類推轉換欄位。

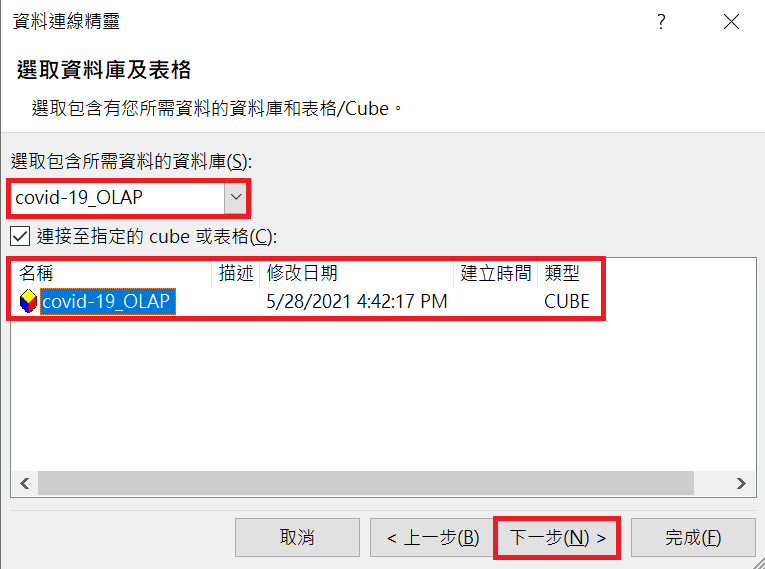
**步驟五**：將轉換好的資料複製到新的資料表中，一份事實表單與數份維度表單，後另存新檔完成資料預處理。

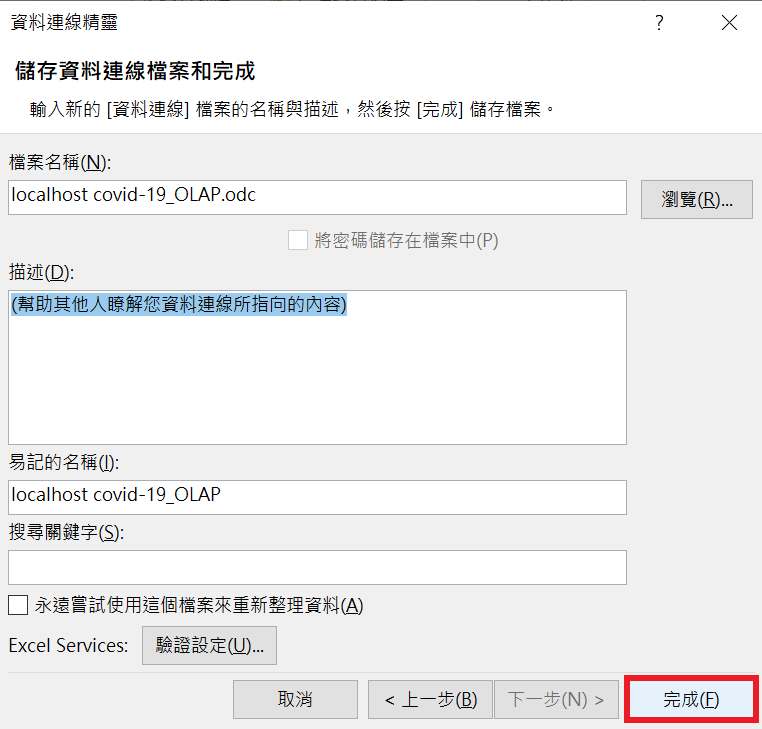
(事實表單) (維度表單)

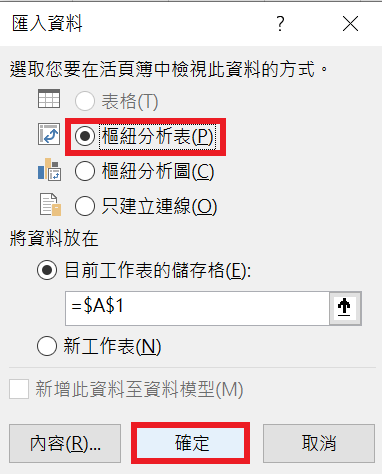
1. 事實表單
2. 維度表單
3. SQL Server分析管理員(資料匯入)
4. 將各個維度表單設定主鍵連結到事實表單
5. 星狀綱要圖Star Schema
6. 表單匯入Visual Studio建立維度與Cube
7. 在Excel中匯入Cube開始OLAP五大分析

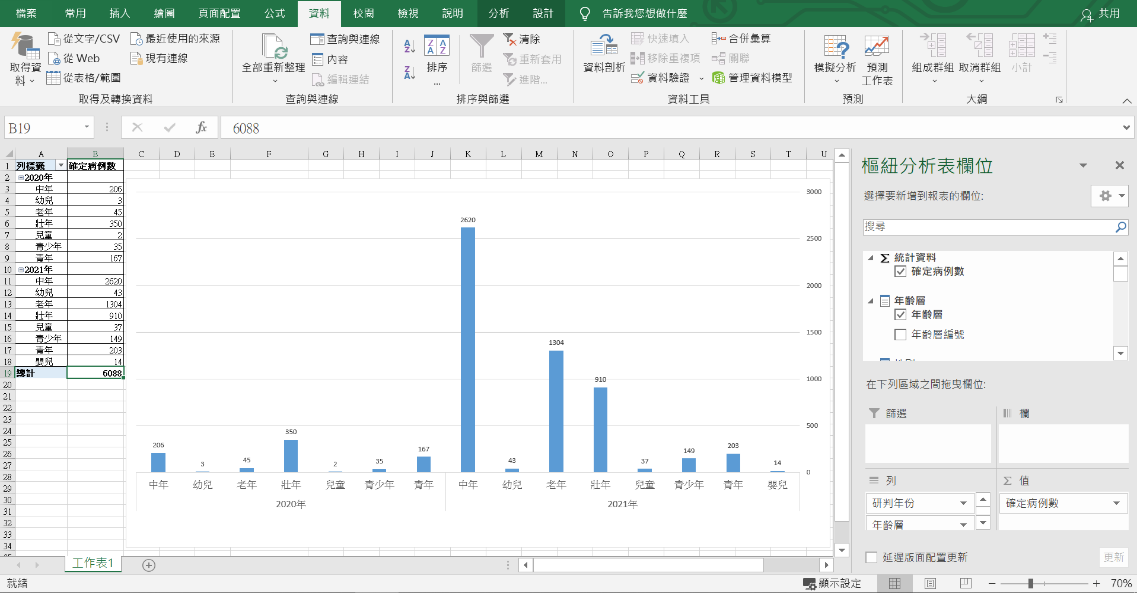
**步驟一**：新增Excel後再資料欄位開啟分析服務連結先前建立的Cube。

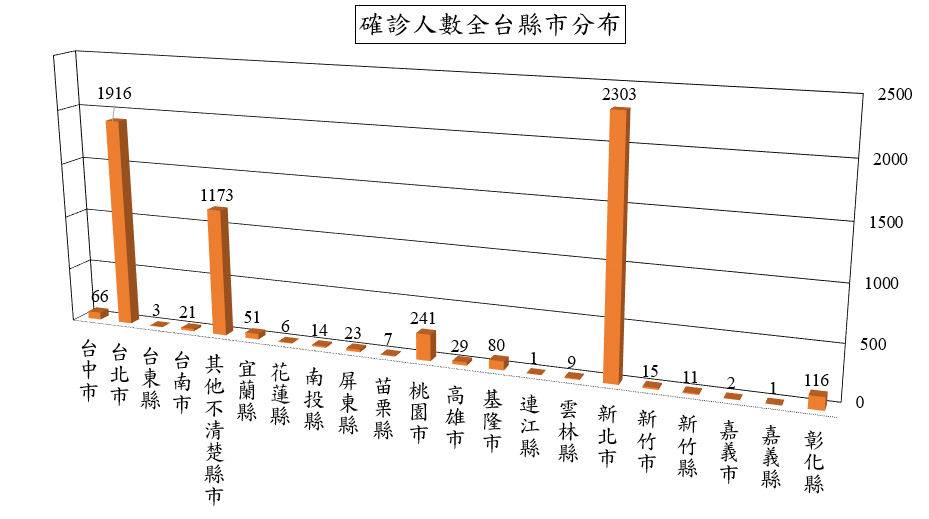
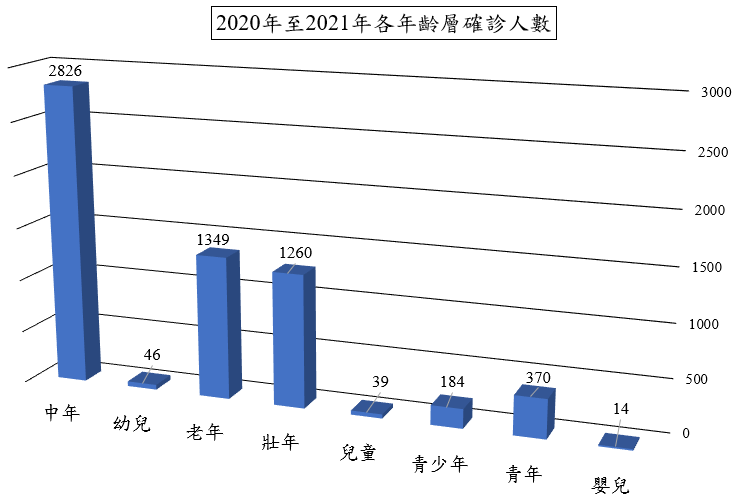
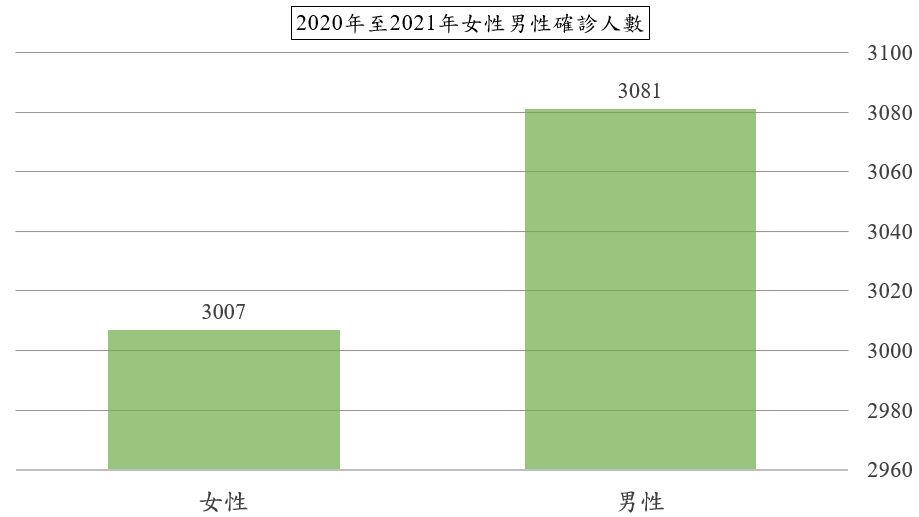
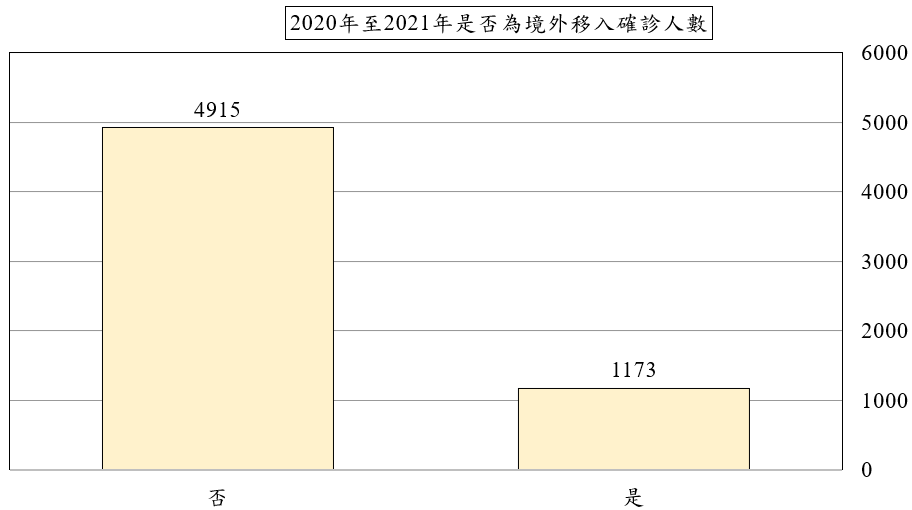
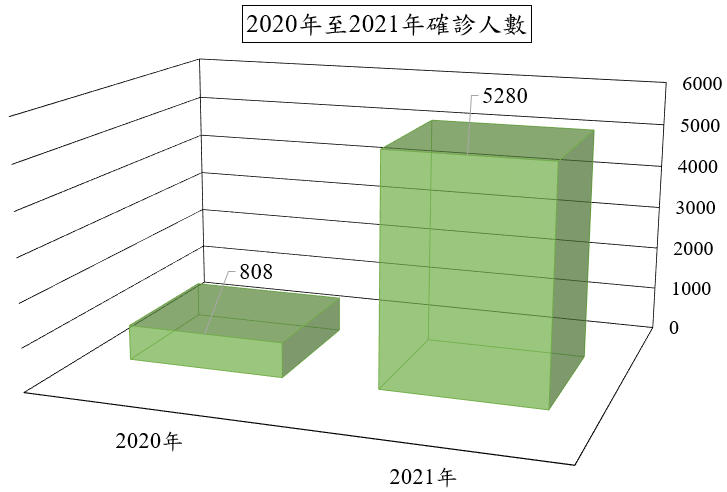
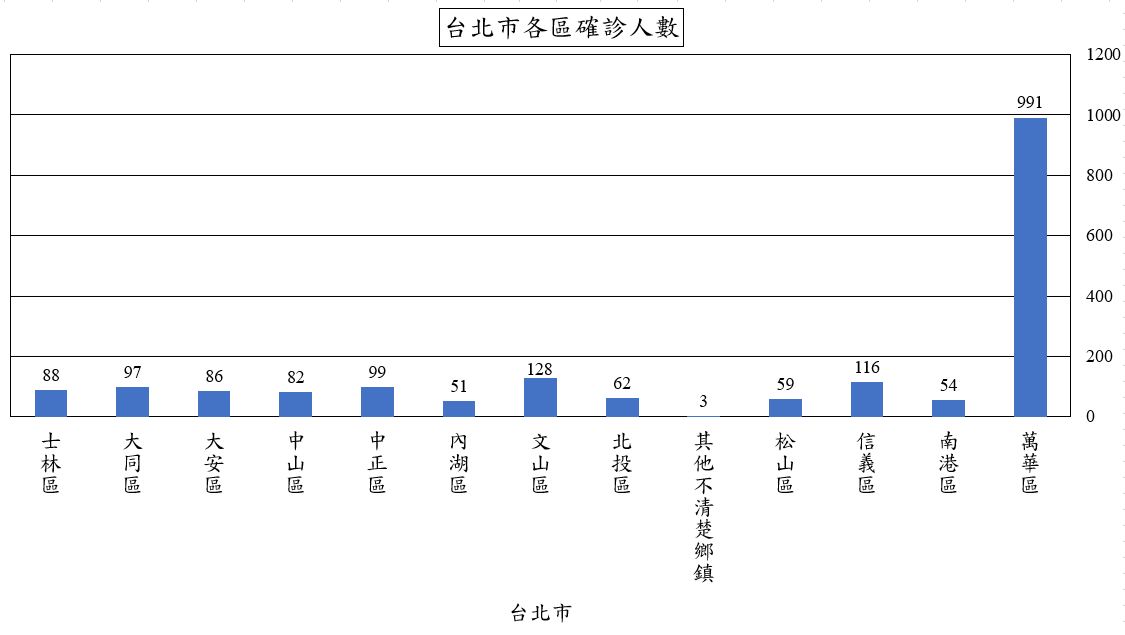
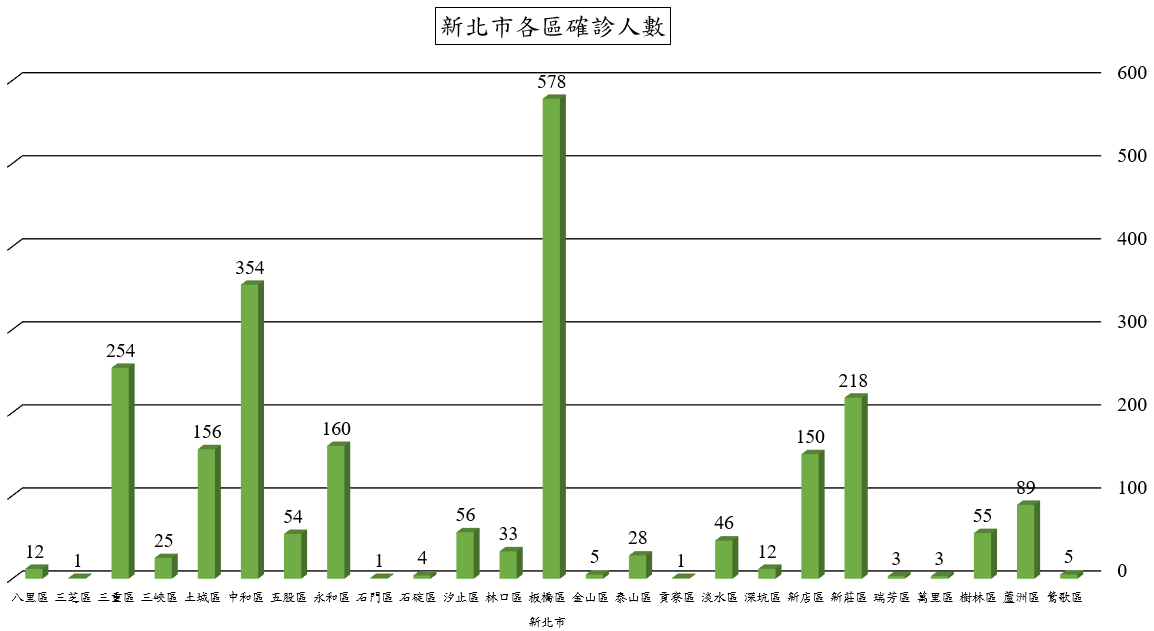
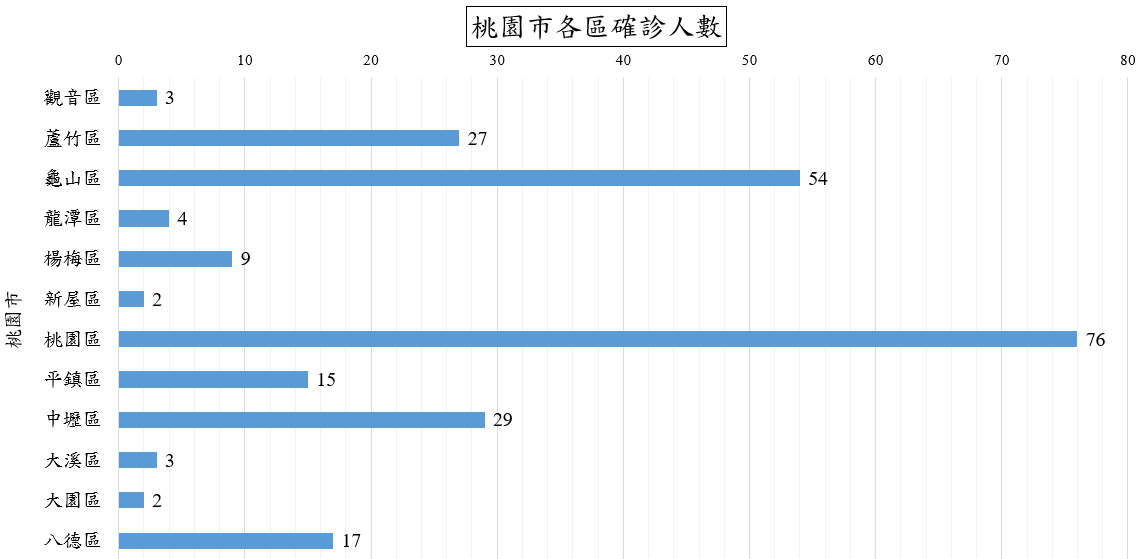
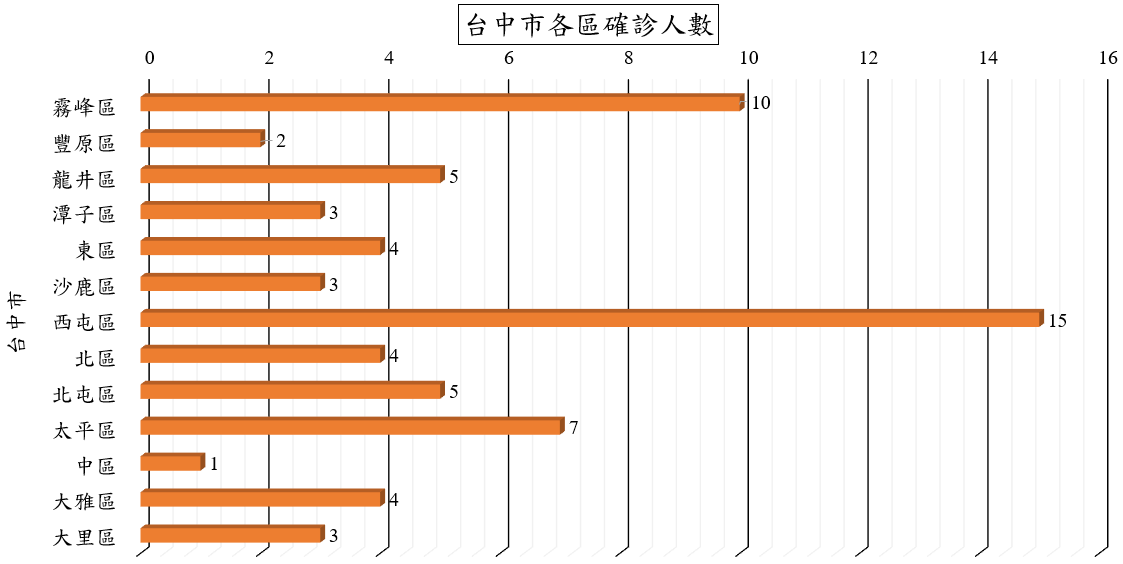
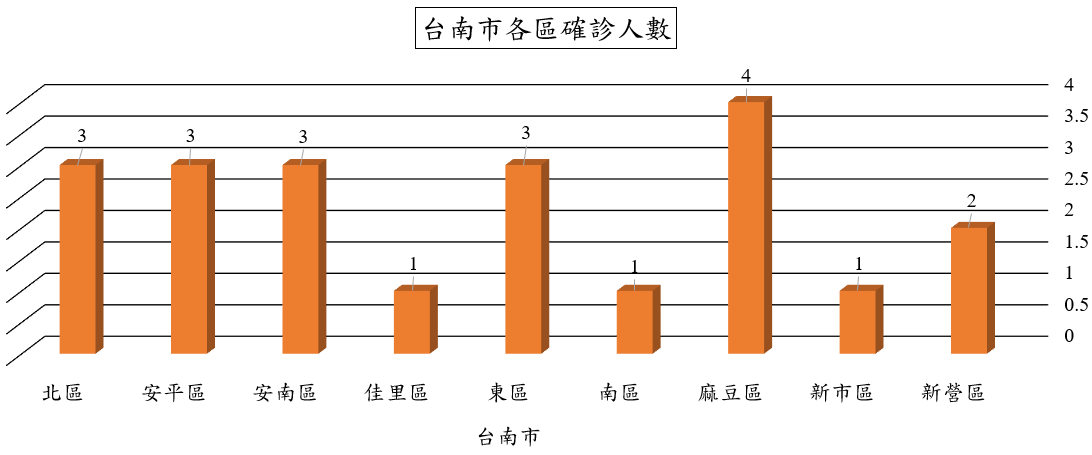
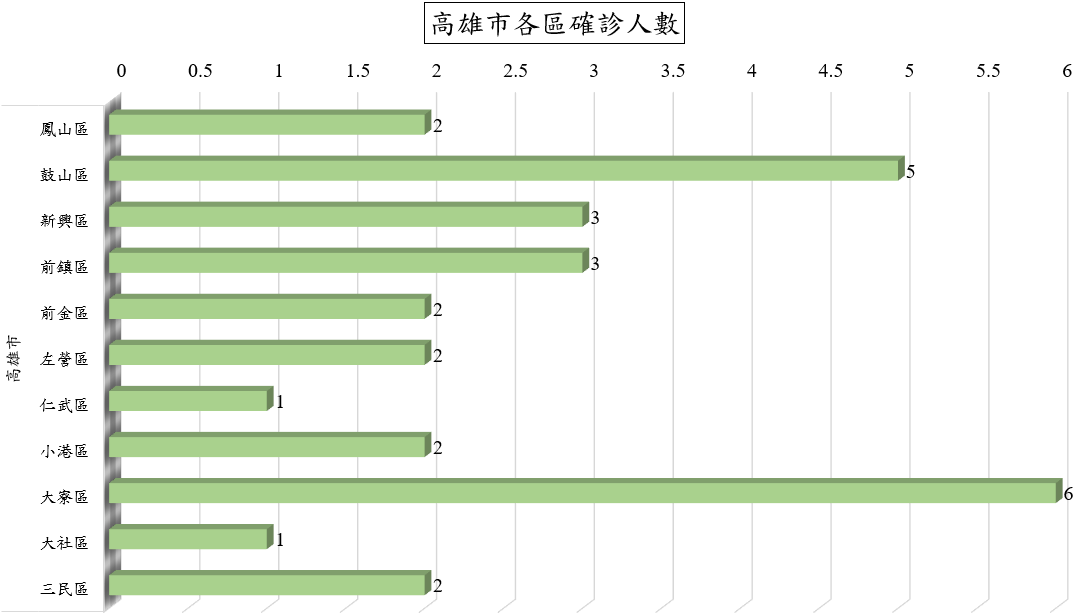
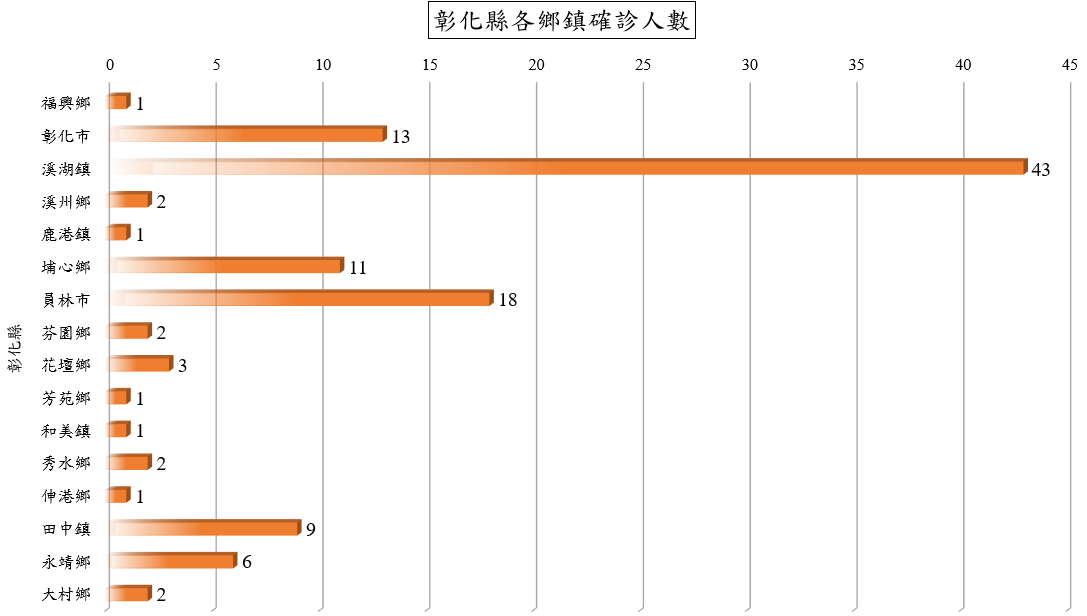
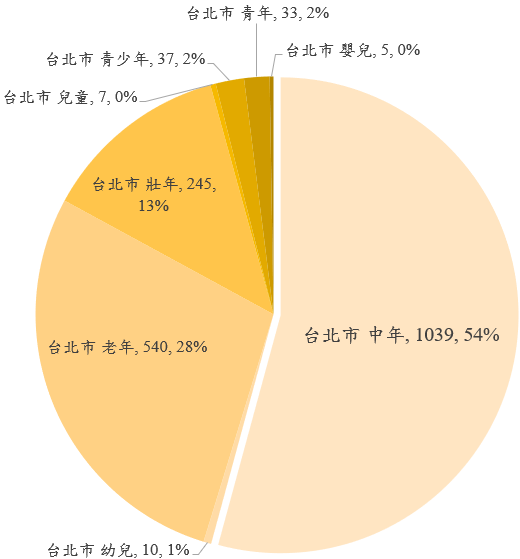
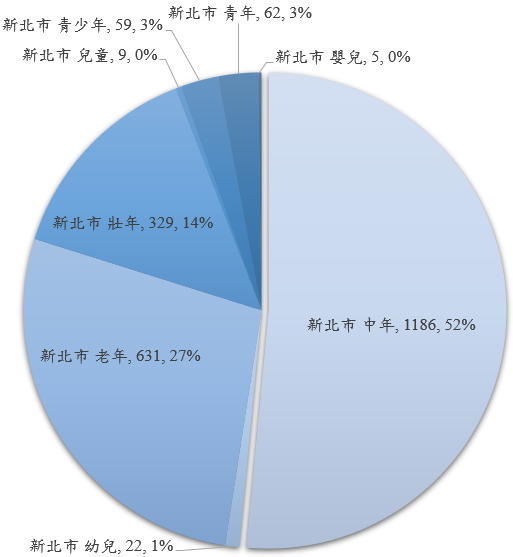
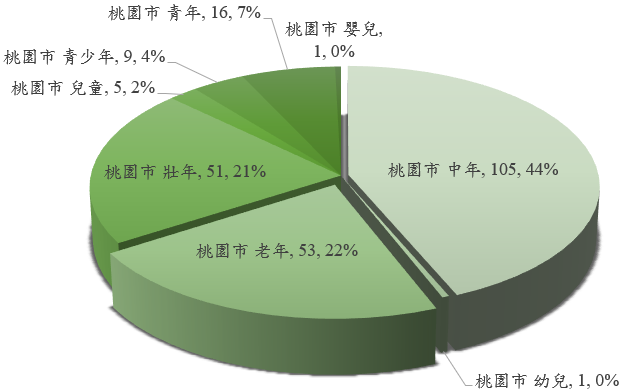
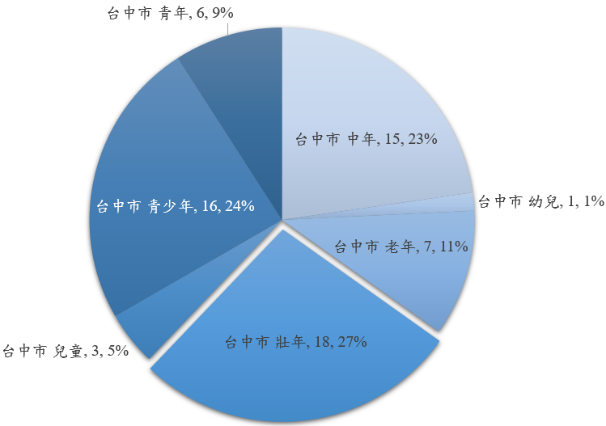
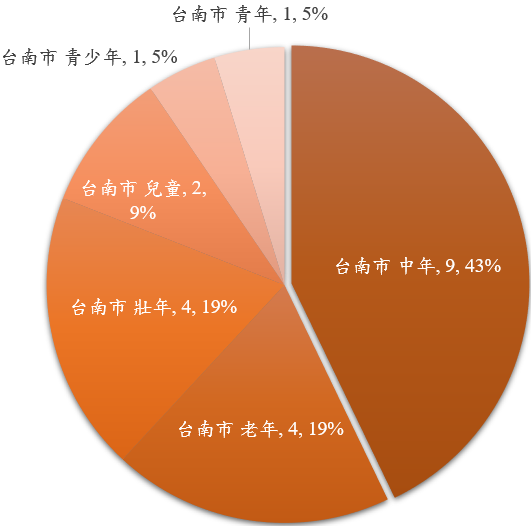
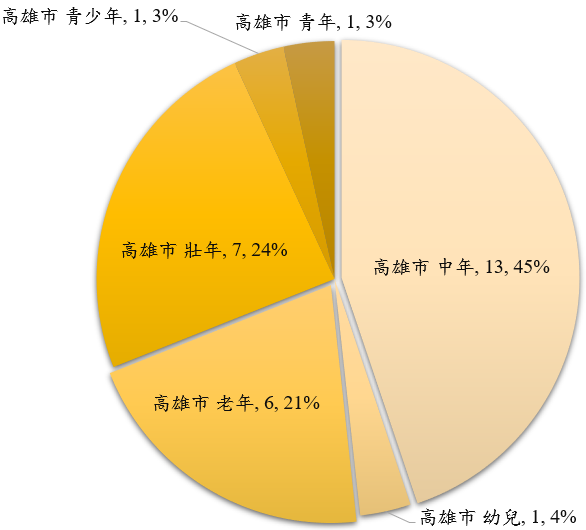
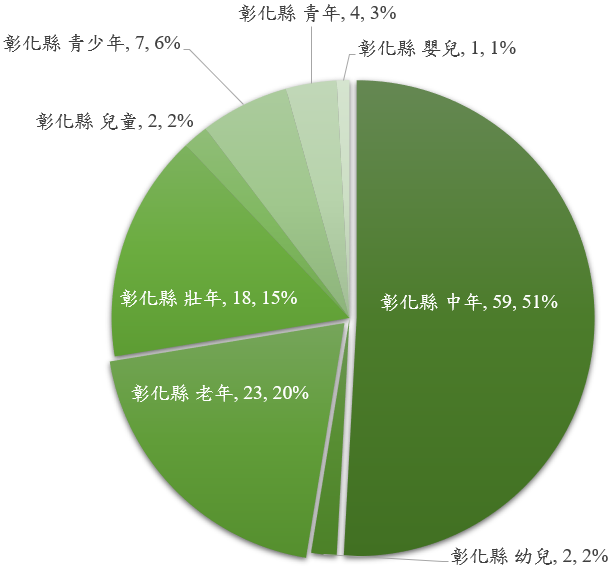
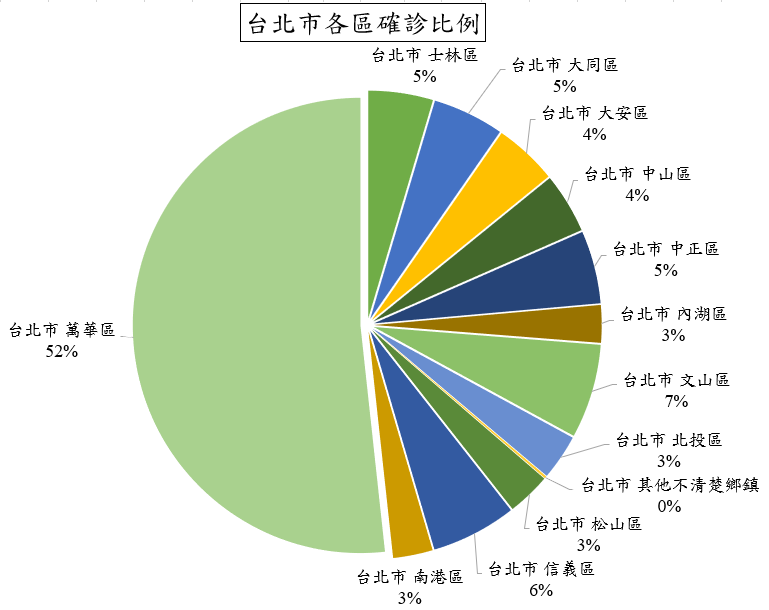
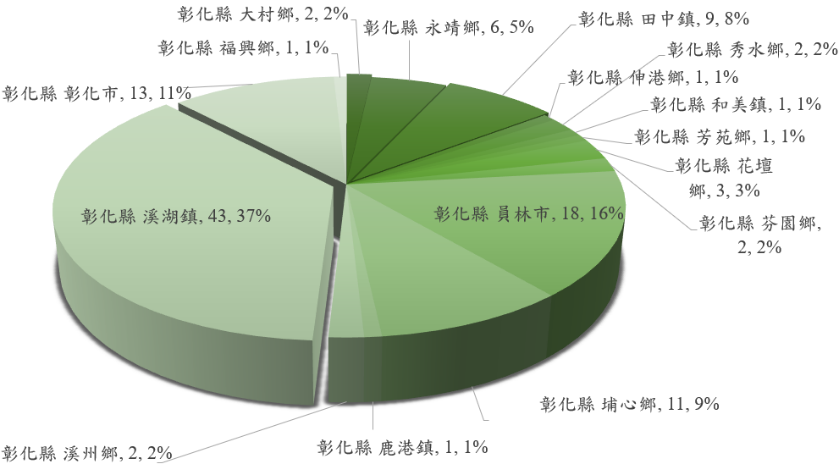
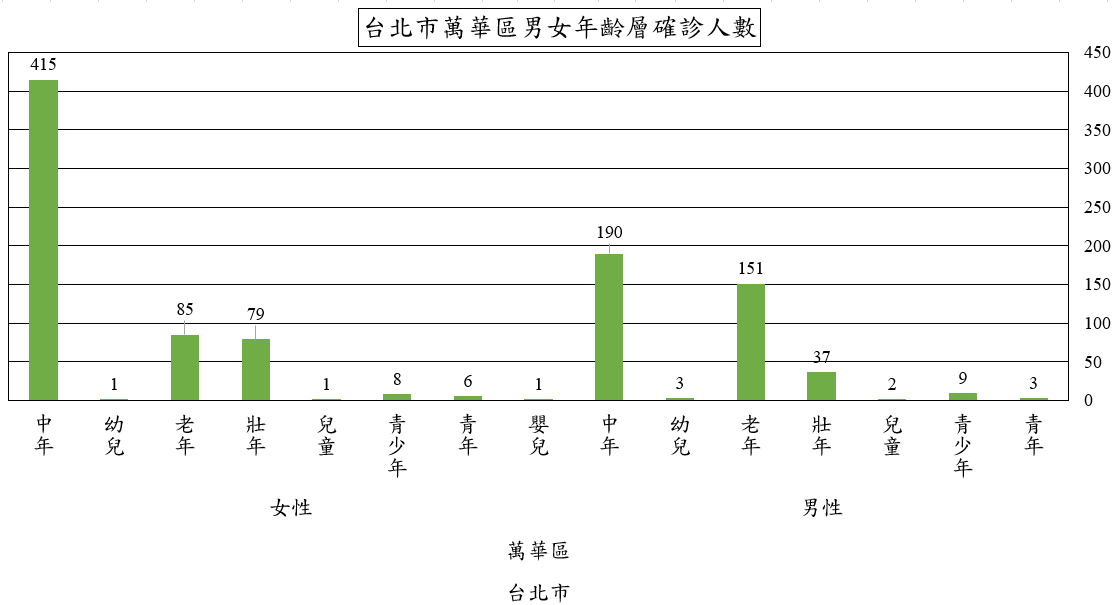
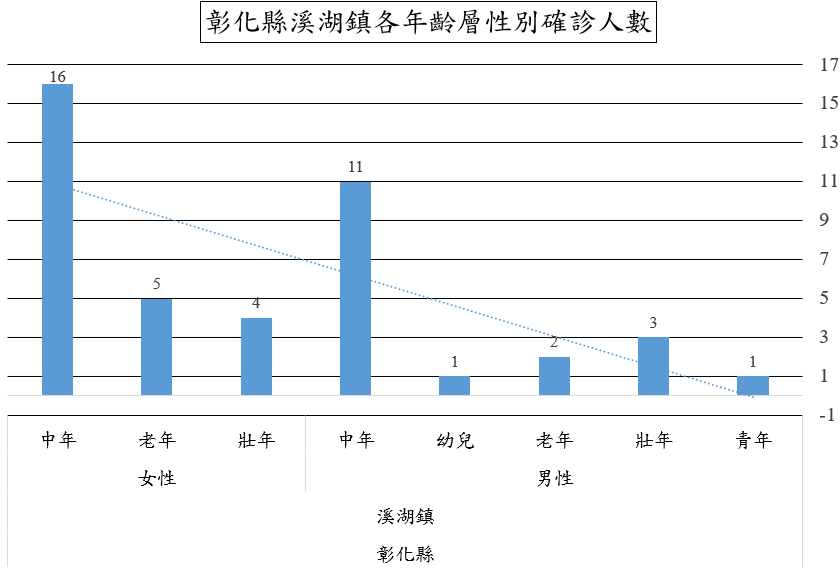
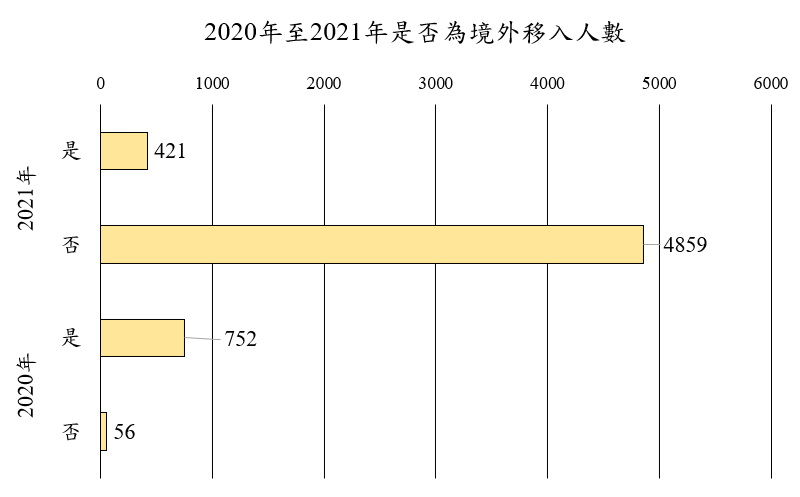
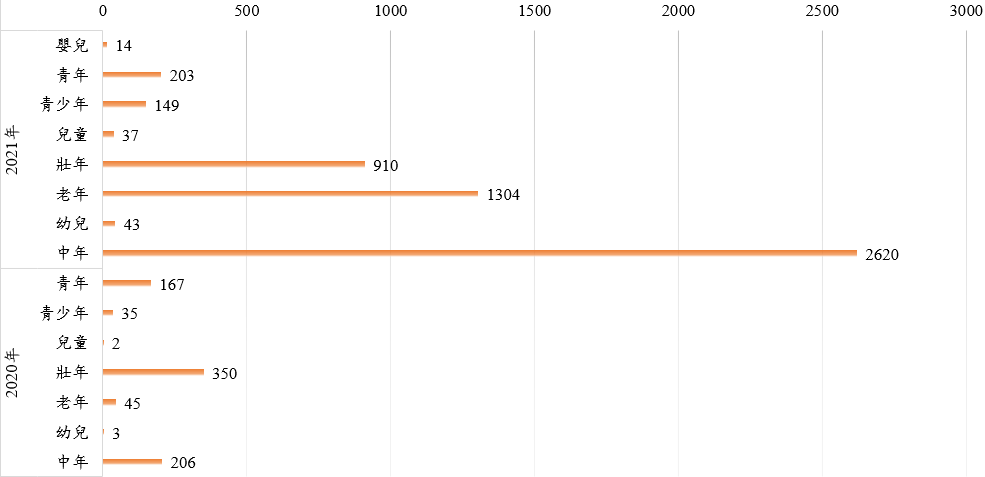
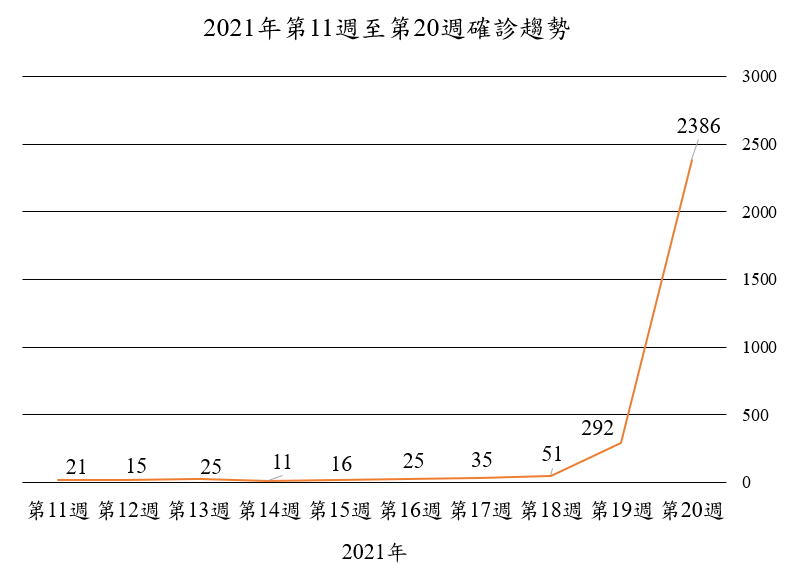
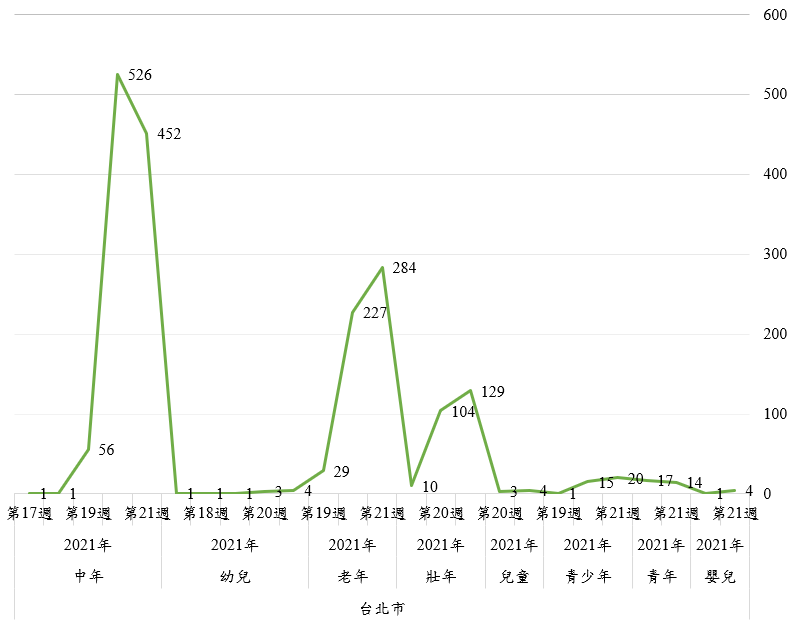
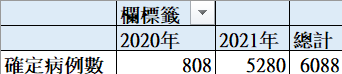
**步驟二**：輸入建立的伺服器名稱localhost連線到資料庫，點選下一步。

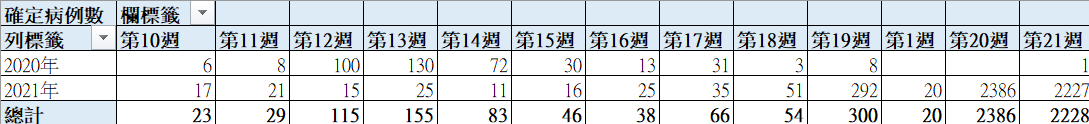
**步驟三**：選取在Visual Studio中建立的Cube後點選下一步。

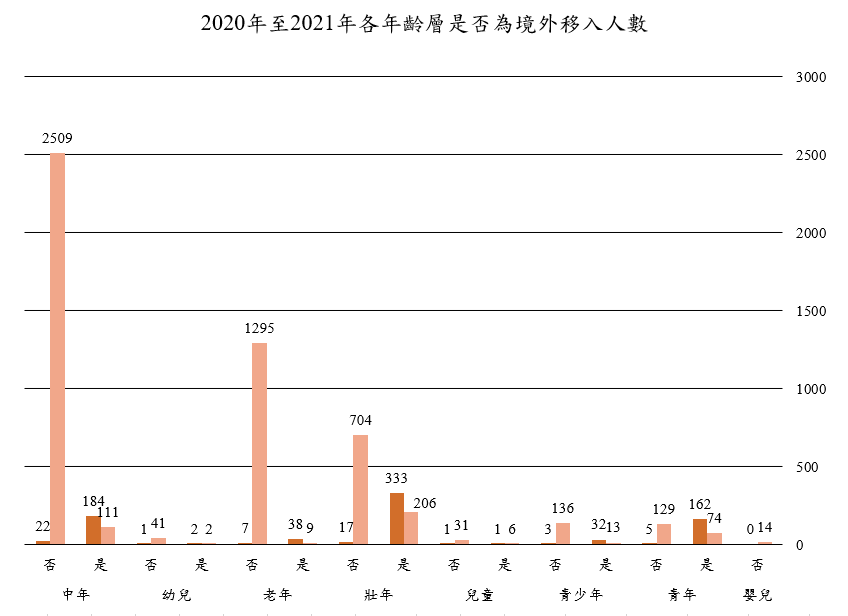
**步驟四**：依序點擊下一步後接著按完成。

**步驟五**：點選樞紐分析表後按確定，即可開始OLAP五大分析作業。

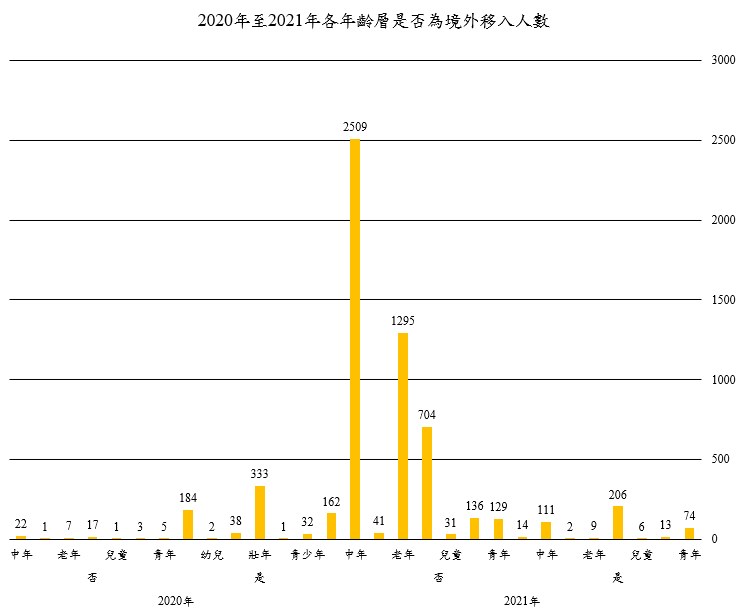
**步驟六**：點選右方的樞紐分析欄位即可變換各資料維度分析其關聯性。

1. 切片(Slice)分析
2. 確診人數全台**縣市**分布(縣市維度)
3. 2020年至2021年各**年齡層**確診人數(年齡層維度)
4. 2020年至2021年**女性男性**確診人數(性別維度)
5. 2020年至2021年**是否為境外移入**確診人數(是否為境外移入維度)
6. **2020年至2021年**確診人數(年份維度)
7. 切丁(Dice)分析
8. 台北市各區確診人數
9. 新北市各區確診人數
10. 桃園市各區確診人數
11. 台中市各區確診人數
12. 台南市各區確診人數
13. 高雄市各區確診人數
14. 彰化縣各鄉鎮確診人數
15. 台北市各年齡層確診人數比例
16. 新北市各年齡層確診人數比例
17. 桃園市各年齡層確診人數比例
18. 台中市各年齡層確診人數比例
19. 台南市各年齡層確診人數比例
20. 高雄市各年齡層確診人數比例
21. 彰化縣各年齡層確診人數比例
22. 台北市各區確診人數比例
23. 彰化縣各鄉鎮確診人數比例
24. 台北市萬華區各年齡層性別確診人數
25. 彰化縣溪湖鎮各年齡層性別確診人數
26. 2020年至2021年是否為境外移入確診人數
27. 2020年至2021年各年齡層確診人數
28. 2021年第11週至第20週確診趨勢
29. 2021年台北市第15週至第21週各年齡層確診人數趨勢圖
30. 向下擷取(Drill-Down)與向上截取(Roll-Up)



1. 轉軸(Pivot)分析
2. 2020年至2021年是否為境外移入確診人數

上圖轉軸成下圖



1. OLAP分析結果
2. 依照縣市染疫人數來看，新北市與台北市位居一、二名，由此可知爆發地點確實都在北部居多，再來就是彰化縣，也是因為爆發市場群聚才有破百例的跡象。
3. 以年齡層分佈來看，中年人與老年人為此次疫情下的最高危險群。
4. 以性別來說，男性確診率高於女性。
5. 非境外人數確診比例提高也都集中在2021年五月份，突然爆發千筆以上人數，且現階段也持續攀升。
6. 以縣市區鄉鎮切丁來看，台北市萬華區與彰化縣溪湖鎮為此次疫情的重災區，以SQL分析結果來看，與衛服部分析結果大致吻合。
7. 2021年第11週至第20週確診趨勢圖明顯遽增，也是從2020年2月份以來，再度有這麼明顯的趨勢圖，也證明國人在疫情上的確鬆懈了。
8. 2020年只有不到千筆的確診人數，但到了2021年卻暴增了5300筆確診病例數，2021年五月份占了最大宗，直到現在還在持續上升當中。
9. 2021年5月的新北市與台北市爆發的新冠肺炎蔓延至全台，造成原本抑制的人數又控制不住，社區感染成為隱憂，無症狀的感染者在不知情的狀態下染疫，成為此次爆發的關鍵點。
10. 心得

**B10756038施宗佑**

經由此次的OLAP分析，讓我更瞭解新冠肺炎的爆發地區，透過疾管署開放資料來分析，著實有參考性，針對不同的維度找出其關聯性，我認為透過OLAP分析有幫助我思考到更深層的資料探勘意義，此次新冠疫情沉寂一年後突然又大爆發，讓全國民眾人心惶惶，地區性群聚為主要引爆點，分別在台北和彰化往全國延伸，而這次的OLAP報告讓我瞭解到資料處理的重要性，資料前處理的過程雖然複雜，但這也是訓練日後進入資訊業會遇到的難題，這份報告花較多的時間在處理雜亂資料，因此，讓我更清楚這份資料集，相信未來能充分利用資料分析技能回饋社會。