

1. (20%) 現有一名為 population.csv 的檔案，內部存著 22 個縣市共 368 個鄉鎮市區的 2018 年公投第十二案結果如右表（僅列出前幾筆資料作為示例，資料總筆數為 368 筆，也就是該 368 個鄉鎮市區）。請撰寫程式以計算匯總出「各縣市」的同意票比例。可用任何一種程式語言，但必須說明要如何讀檔，以及讀取檔案後所獲得的變數型態。（讀檔方法與變數型態說明 5%、匯總部分的程式碼 15%）。

Abc Extract 公投案號	Abc Extract 縣市	Abc Extract 鄉鎮市區	# Extract 同意票數	# Extract 不同意票數
Ref12	連江縣	南竿鄉	2,094	1,064
Ref12	連江縣	北竿鄉	713	347
Ref12	連江縣	莒光鄉	467	289
Ref12	連江縣	東引鄉	325	225
Ref12	金門縣	金城鎮	8,177	4,455
Ref12	金門縣	金沙鎮	4,266	1,966
Ref12	金門縣	金湖鎮	5,863	3,118
Ref12	金門縣	金寧鄉	5,802	3,895
Ref12	金門縣	烈嶼鄉	2,711	1,293
Ref12	金門縣	烏坵鄉	113	46
Ref12	宜蘭縣	宜蘭市	25,673	17,297
Ref12	宜蘭縣	羅東鎮	18,673	12,875
Ref12	宜蘭縣	蘇澳鎮	10,125	7,865
Ref12	宜蘭縣	頭城鎮	7,489	5,317
Ref12	宜蘭縣	礁溪鄉	9,089	6,782
Ref12	宜蘭縣	壯圍鄉	6,976	4,687

2. (25%) 「誘餌式標題 (Clickbait)」泛指媒體設計來讓閱聽人點閱，賺取相對應的廣告收益的標題，如「你不可不知道的十件事」便應用整理資訊的方法來引誘點擊。請舉出一種你曾經看過，以誘騙點擊率為目的的標題寫法二則 (5%)。請用任一種程式來設計一個函式，目的是抽取你所觀察到的誘餌式標題，函式輸入是一個新聞標題句子，函式回傳值應該為一個數值或 T/F (15%)。今有一新聞標題資料庫共 3000 則，每則有該新聞的標題、時間、報紙來源等。我想要找人來「標記」每則新聞的誘餌性，請問要如何確保標記結果的信效度 (5%)？

3. (15%) 在做數值呈現時，我們經常要將「1, 23, 456, 7890」等數字加上前綴零 (Leading zero) 成為「0001, 0023, 0456, 7890」的型態，請寫一個函式，偵測最大數字後，把比他小的數字自動視需要補前綴零，例如有一筆資料為「3, 45, 456789」，則輸出為「000003, 000045, 456789」(15%)。

4. (20%) 簡答 (每題 5%)

- 1) 請簡要描述 Docker 為何？有何優點？
- 2) 請寫出一個 SQL 指令，從一資料表 mytable 選擇兩個變項 var1, var2、並篩選 var1 值大於 100 的資料項。
- 3) 列舉至少三個你曾聽過的資料庫管理系統；說明 NoSQL 的優缺點為何？
- 4) 資料庫管理系統的 Primary key 與 foreign key 是什麼？如何使用？

5. (20%) 給定一篇英文文章，寫一個函式，計算每個字的出現次數，並按照出現次數的多寡由大而小排序 (20%，可用熟悉的程式語言撰寫，如 R、Python、JavaScript、Java、C++/C 等，但勿夾雜)。