

TQC+ 網頁資料擷取與分析

考前衝衝衝



國立宜蘭大學資訊工程系
吳政瑋 專案助理教授

wucw@niu.edu.tw



101 文化部展覽資訊

- 設計說明：

- 請撰寫一程式，讀取文化部展覽資訊read.json，請將其中title(活動名稱)、showUnit(演出單位)、startDate(活動起始日期)、endDate(活動結束日期)等四個欄位內容轉存為write.csv(需為UTF-8編碼格式)，各欄位內容之間以一個半形逗號隔開。
- 提示：只需要輸出資料，不需要輸出欄位名稱。

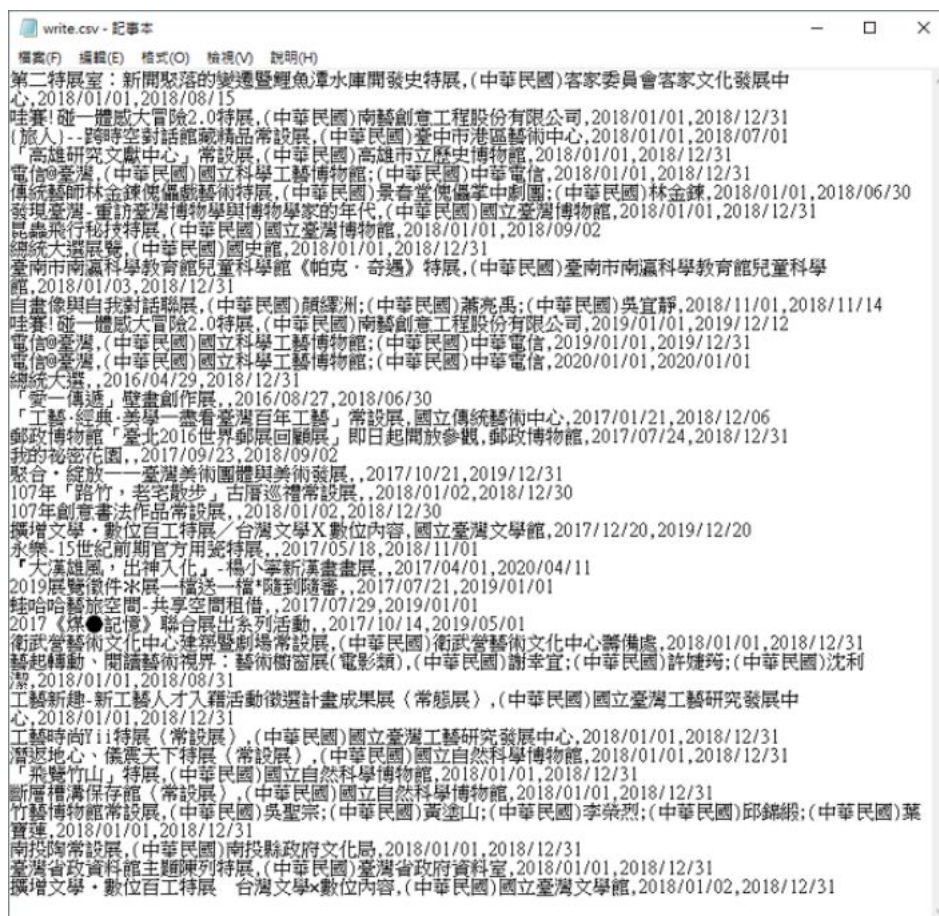


101 文化部展覽資訊

- 輸入說明：
讀取read.json
- 輸出說明：
將四個欄位的內容:title、showUnit、startDate、endDate，輸出至write.csv檔案，各欄位內容之間以一個半形逗號隔開

101 文化部展覽資訊

- 範例輸入:無
- 範例輸出:





102 新北市公共自行車即時資訊

- 設計說明：

- 請撰寫一程式，讀取新北市公共自行車即時資訊 `read.xml`，請將其中 `sno` (站點代號)、`sna` (中文場站名稱)、`tot` (場站總停車格) 等三個欄位轉存為 `write.csv` (需為UTF-8編碼格式)，各欄位內容之間以一個半形逗號隔開。
- 提示：只需要輸出資料，不需要輸出欄位名稱。

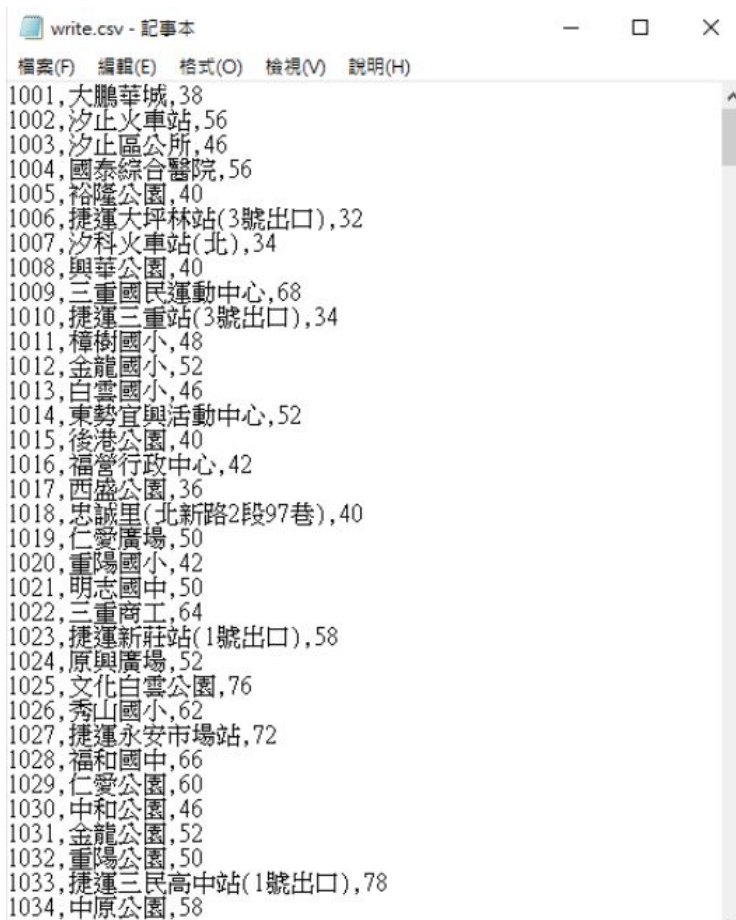


102 新北市公共自行車即時資訊

- 輸入說明：
讀取read.xml
- 輸出說明：
將三個欄位的內容:sno、sna、tot, 輸出write.csv
檔案, 各欄位內容之間以一個半形逗號隔開

102 新北市公共自行車即時資訊

- 範例輸入:無
- 範例輸出:



```
write.csv - 記事本
檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)
1001,大鵬華城,38
1002,汐止火車站,56
1003,汐止區公所,46
1004,國泰綜合醫院,56
1005,裕隆公園,40
1006,捷運大坪林站(3號出口),32
1007,汐科火車站(北),34
1008,興華公園,40
1009,三重國民運動中心,68
1010,捷運三重站(3號出口),34
1011,樟樹國小,48
1012,金龍國小,52
1013,白雲國小,46
1014,東勢宜興活動中心,52
1015,後港公園,40
1016,福營行政中心,42
1017,西盛公園,36
1018,忠誠里(北新路2段97巷),40
1019,仁愛廣場,50
1020,重陽國小,42
1021,明志國中,50
1022,三重商工,64
1023,捷運新莊站(1號出口),58
1024,原興廣場,52
1025,文化白雲公園,76
1026,秀山國小,62
1027,捷運永安市場站,72
1028,福和國中,66
1029,仁愛公園,60
1030,中和公園,46
1031,金龍公園,52
1032,重陽公園,50
1033,捷運三民高中站(1號出口),78
1034,中原公園,58
```



103 勞保投保薪資分級表

- 設計說明：

- 請撰寫一程式，讀取勞保投保薪資分級表read.json內的資料後，將資料轉存為write.yaml。



103 勞保投保薪資分級表

- 輸入說明：
讀取read.json
- 輸出說明：
將資料輸出至write.yaml

103 勞保投保薪資分級表

- 範例輸入:無
- 範例輸出:



- 投保薪資等級: 第1級
月投保薪資: 22000元
月薪資總額: 22000元以下
身分別: 一般勞工
- 投保薪資等級: 第2級
月投保薪資: 22800元
月薪資總額: 22001元至22800元
身分別: 一般勞工
- 投保薪資等級: 第3級
月投保薪資: 24000元
月薪資總額: 22801元至24000元
身分別: 一般勞工
- 投保薪資等級: 第4級
月投保薪資: 25200元
月薪資總額: 24001元至25200元
身分別: 一般勞工
- 投保薪資等級: 第5級
月投保薪資: 26400元
月薪資總額: 25201元至26400元
身分別: 一般勞工
- 投保薪資等級: 第6級
月投保薪資: 27600元
月薪資總額: 26401元至27600元
身分別: 一般勞工
- 投保薪資等級: 第7級
月投保薪資: 28800元
月薪資總額: 27601元至28800元
身分別: 一般勞工



104 JSON檔案輸出處理

- 設計說明：
 - 請撰寫一程式，建立以下資料並將其輸出為write.json檔案：

```
{
  'people' :
  [{
    'id': '1',
    'name': 'Peter',
    'country': 'Taiwan'
  },
  {
    'id': '2',
    'name': 'Jack',
    'country': 'USA'
  },
  {
    'id': '3',
    'name': 'Cindy',
    'country': 'Japan'
  }
  ]
}
```



104 JSON檔案輸出處理

- 輸入說明：
無
- 輸出說明：
將資料輸出至write.json

104 JSON檔案輸出處理

- 範例輸入: 無
- 範例輸出:



```
{ "people": [ { "id": "1", "name": "Peter",  
"country": "Taiwan" }, { "id": "2", "name":  
"Jack", "country": "USA" }, { "id": "3",  
"name": "Cindy", "country": "Japan" } ] }
```



105 受僱員工資料表

- 設計說明：
 - 請撰寫一程式連結read.db資料庫，讀取Employee資料表，輸出Employee資料表內的資料。



105 受僱員工資料表

- 輸入說明：
無
- 輸出說明：
輸出Employee資料表內的資料



105 受僱員工資料表

- 範例輸入: 無
- 範例輸出:

```
(1, '小陳', 1997, '新北市', 58000)
(2, '小范', 2000, '臺北市', 50000)
(3, '小施', 1999, '高雄市', 47000)
(4, '小吳', 1998, '台中市', 52000)
```




201 搜尋字詞

- 設計說明：

- 請撰寫一程式, 爬取

<http://tqc.codejudge.com:3000/target/5201.html>, 程式須回傳下列資訊: 讓使用者輸入欲搜尋的字詞, 再輸出字詞的搜尋結果及字詞出現的次數。



201 搜尋字詞

- 輸入說明：
爬取網頁
搜尋的字詞
- 輸出說明：
字詞的搜尋結果
字詞出現的次數



201 搜尋字詞

- 範例輸入：無
- 範例輸出：

```
請輸入欲搜尋的字串  : TQC+  
TQC+ 搜尋成功  
TQC+ 出現 23 次
```



202 美元收盤匯率

- 設計說明：

- 請撰寫一程式，爬取read.html，取得「新臺幣對美元銀行間成交之收盤匯率」資料，並將其中日期、NTD/USD兩個欄位的名稱與資料轉存為write.csv (需為UTF-8編碼格式)。



202 美元收盤匯率

- 輸入說明：
爬取網頁
- 輸出說明：
日期、NTD/USD兩個欄位的名稱與資料，輸出至
write.csv

202 美元收盤匯率

- 範例輸入:無
- 範例輸出:



日期,NTD/USD
2018-06-26,30.412
2018-06-25,30.403
2018-06-22,30.307
2018-06-21,30.302
2018-06-20,30.170
2018-06-19,30.186
2018-06-15,30.002
2018-06-14,29.938
2018-06-13,29.890
2018-06-12,29.860
2018-06-11,29.821
2018-06-08,29.816
2018-06-07,29.740
2018-06-06,29.738
2018-06-05,29.815
2018-06-04,29.820
2018-06-01,29.864
2018-05-31,29.980
2018-05-30,30.050
2018-05-29,29.988
2018-05-28,29.928
2018-05-25,29.945



203 台灣彩券

- 設計說明：

- 請撰寫一程式, 爬取

<http://tqc.codejudger.com:3000/target/5203.html>, 程式須回傳下列資訊：

- 大樂透的開出順序
 - 大樂透的大小順序
 - 大樂透的特別號



203 台灣彩券

- 輸入說明：
爬取網頁
- 輸出說明：
 - 大樂透的開出順序
 - 大樂透的大小順序
 - 大樂透的特別號



203 台灣彩券

- 範例輸入：無
- 範例輸出：

大樂透開獎：

開出順序：	44	25	18	14	16	23
大小順序：	14	16	18	23	25	44
特別號：	41					



204 新北市大專院校名單

- 設計說明：

- (1) 請撰寫一程式，爬取新北市大專院校名單，API連結如下
: `http://tqc.codejudge.com:3000/target/5204.json`
- (2) 程式須輸出：新北市每一所大專院校的相關訊息：名稱、地址、聯絡電話、網站、資料更新時間。



204 新北市大專院校名單

- 輸入說明：
爬取API資料
- 輸出說明：
新北市每一所大專院校的相關訊息：名稱、地址、
聯絡電話、網站、資料更新時間



204 新北市大專院校名單

- 範例輸入：無
- 範例輸出：

新北市大專院校名單：

名稱：馬偕醫專三芝校區

地址：新北市三芝區中正路三段42號

聯絡電話：02-26366799

網站：www.mkc.edu.tw

資料更新時間：2018-09-12 06:00:01.0

名稱：馬偕醫學院

地址：新北市三芝區中正路三段46號

聯絡電話：02-26360303

網站：www.mmc.edu.tw

資料更新時間：2018-09-12 06:00:01.0

名稱：法鼓大學

地址：新北市金山區

聯絡電話：

網站：www.ddc.edu.tw/zh-tw

資料更新時間：2018-09-12 06:00:01.0



205 空氣品質指標 (AQI)

- 設計說明：

- (1) 請撰寫一程式, 爬取政府AQI開放資料, API連結:
<http://tqc.codejudger.com:3000/target/5205.json>
- (2) 程式須回傳下列資訊:
- 內容長度
- 新北市每一個地區的相關訊息: 地區名稱、AQI指數、PM2.5指數、PM10指數、資料更新時間;
- 在輸出時, AQI指數、PM2.5指數、PM10指數與資料更新時間四項資訊前加入一個 tab 鍵(\t)



205 空氣品質指標 (AQI)

- 輸入說明：
爬取API資料
- 輸出說明：
 - 內容長度
 - 新北市每一個地區的相關訊息：地區名稱、AQI指數、PM2.5指數、PM10指數、資料更新時間
 - 在輸出時，AQI指數、PM2.5指數、PM10指數與資料更新時間四項資訊前加入一個 tab 鍵(\t)

205 空氣品質指標 (AQI)

- 範例輸入: 無
- 範例輸出:

Content-Length: 37985

新北市PM2.5相關資料:

汐止:

AQI : 36
PM2.5 : 16
PM10 : 36
資料更新時間: 2018-06-27 13:00

萬里:

AQI : 49
PM2.5 : 10
PM10 : 20
資料更新時間: 2018-06-27 13:00

新店:

AQI : 45
PM2.5 : 22
PM10 : 22
資料更新時間: 2018-06-27 13:00

土城:

AQI : 60
PM2.5 : 20
PM10 : 32
資料更新時間: 2018-06-27 13:00

板橋:

AQI : 38
PM2.5 : 15
PM10 : 42
資料更新時間: 2018-06-27 13:00

新莊:

AQI : 42
PM2.5 : 12
PM10 : 27
資料更新時間: 2018-06-27 13:00

菜寮:

AQI : 60
PM2.5 : 20
PM10 : 32
資料更新時間: 2018-06-27 13:00

林口:

AQI : 58
PM2.5 : 19
PM10 : 27
資料更新時間: 2018-06-27 13:00

淡水:

AQI : 40
PM2.5 : 10
PM10 : 16
資料更新時間: 2018-06-27 13:00

三重:

AQI : 60
PM2.5 : 19
PM10 : 62
資料更新時間: 2018-06-27 13:00

永和:

AQI : 50
PM2.5 : 21
PM10 : 35
資料更新時間: 2018-06-27 13:00

富貴角:

AQI : 42
PM2.5 :
PM10 : 11
資料更新時間: 2018-06-27 13:00



301 學生成績

- 設計說明：
 - 利用程式內提供的數據，依下列要求進行輸出：
 - 輸出全部學生的成績，行標題設為科目的名稱；列標題為每個人的名字
 - 輸出後二位學生的所有成績
 - 將自然成績做遞減排序輸出
 - 僅列小黃的成績，並將其英文成績改為80



301 學生成績

- 輸入說明：
無
- 輸出說明：
 - 輸出全部學生的成績，行標題設為科目的名稱；列標題為每個人的名字
 - 輸出後二位學生的所有成績
 - 將自然成績做遞減排序輸出
 - 僅列小黃的成績，並將其英文成績改為80

301 學生成績

- 範例輸入: 無
- 範例輸出:

行標題為科目，列標題為個人的所有學生成績

	國語	數學	英文	自然	社會
小林	75	62	85	73	60
小黃	91	53	56	63	65
小陳	71	88	51	69	87
小美	69	53	87	74	70

後二位的成績

	國語	數學	英文	自然	社會
小陳	71	88	51	69	87
小美	69	53	87	74	70

以自然遞減排序

小美	74
小林	73
小陳	69
小黃	63

Name: 自然, dtype: int64

小黃的成績

國語	91
數學	53
英文	80
自然	63
社會	65

Name: 小黃, dtype: int64



302 矩陣

- 設計說明：
 - 請用numpy隨機產生5~15之間, 15個正整數並輸出
 - 請將 1. 轉成 3×5 的X矩陣並輸出
 - 請輸出X矩陣的最大值
 - 請輸出X矩陣的最小值
 - 請輸出X矩陣的總和
 - 請輸出X矩陣四個角落的元素內容



302 矩陣

- 輸入說明：
無
- 輸出說明：
 - 請用numpy隨機產生5~15之間, 15個正整數並輸出
 - 請將 1. 轉成 3×5 的X矩陣並輸出
 - 請輸出X矩陣的最大值
 - 請輸出X矩陣的最小值
 - 請輸出X矩陣的總和
 - 請輸出X矩陣四個角落的元素內容



302 矩陣

- 範例輸入：無
- 範例輸出：

```
隨機正整數： [ 7  7 11  6  8 15 14 11  6  5  6 14  5  5 14]
```

```
X矩陣內容：
```

```
[[ 7  7 11  6  8]  
 [15 14 11  6  5]  
 [ 6 14  5  5 14]]
```

```
最大： 15
```

```
最小： 5
```

```
總和： 134
```

```
四個角落元素：
```

```
[[ 7  8]  
 [ 6 14]]
```

303 果菜批發市場拍賣行情

- 設計說明：

- 某日各果菜批發市場之西瓜與香瓜之拍賣行情價量表如下：

	西瓜價		西瓜量		香瓜價		香瓜量	
三重區	9		203674		13.2		18894	
台中市	11.7		180785		12.3		54894	
台北一	10.1		127802		14.7		18563	
台北二	11.8		28604		14.9		21963	
台東市	13.2		600		13.1		900	
板橋區	6.9		38071		9.6		3555	
高雄市	12.1		35660		10.6		9005	
嘉義市	12		15000		13		12000	
鳳山區	11.7		48770		9.1		14370	
豐原區	9.84		6100		11.89		8980	



303 果菜批發市場拍賣行情

- 輸入說明：

無

- 輸出說明：

分別輸出下列4段資料：

a. 拍賣行情價量表

b. 以西瓜價遞減排序後，各市場的西瓜價

c. 台北一市場西瓜/香瓜價量的行情

d. 將「三重市」改為「三重區」、「香瓜價」改為「洋香瓜價」、「香瓜量」改為「洋香瓜量」，重新輸出整個表格

303 果菜批發市場拍賣行情

■ 範例輸入: 無

■ 範例輸出:

西瓜與香瓜之拍賣行情價量表

	西瓜價	西瓜量	香瓜價	香瓜量
三重市	9.00	203674	13.20	18894
台中市	11.70	180785	12.30	54894
台北一	10.10	127802	14.70	18563
台北二	11.80	28604	14.90	21963
台東市	13.20	600	13.10	900
板橋區	6.90	38071	9.60	3555
高雄市	12.10	35660	10.60	9005
嘉義市	12.00	15000	13.00	12000
鳳山區	11.70	48770	9.10	14370
豐原區	9.84	6100	11.89	8980

台北一市場的行情

西瓜價	10.1
西瓜量	127802.0
香瓜價	14.7
香瓜量	18563.0

Name: 台北一, dtype: float64

以西瓜價遞減排序

台東市	13.20
高雄市	12.10
嘉義市	12.00
台北二	11.80
台中市	11.70
鳳山區	11.70
台北一	10.10
豐原區	9.84
三重市	9.00
板橋區	6.90

Name: 西瓜價, dtype: float64

全體市場行情

	西瓜價	西瓜量	洋香瓜價	洋香瓜量
三重區	9.00	203674	13.20	18894
台中市	11.70	180785	12.30	54894
台北一	10.10	127802	14.70	18563
台北二	11.80	28604	14.90	21963
台東市	13.20	600	13.10	900
板橋區	6.90	38071	9.60	3555
高雄市	12.10	35660	10.60	9005
嘉義市	12.00	15000	13.00	12000
鳳山區	11.70	48770	9.10	14370
豐原區	9.84	6100	11.89	8980



304 資料處理與分析

- 設計說明：
 - 請讀取read.csv中的資料轉換成numpy陣列，並輸出以下資訊：
 - 資料集型態
 - 平均數
 - 中位數
 - 標準差
 - 變異數
 - 極差值
- 註：數值需四捨五入至小數點後兩位



304 資料處理與分析

- 輸入說明：
讀取read.csv的內容
- 輸出說明：
 - 資料集型態
 - 平均數
 - 中位數
 - 標準差
 - 變異數
 - 極差值



304 資料處理與分析

- 範例輸入：無
- 範例輸出：

```
資料型態：<class 'numpy.ndarray'>  
平均值：50.48  
中位數：53.00  
標準差：27.57  
變異數：760.27  
極差值：94.00
```



305 登革熱病例統計

- 設計說明：

請撰寫一程式，讀取登革熱近12個月每日確定病例統計read.csv，接著依序輸出下列項目：

- a. 以遞減順序顯示居住縣市的病例人數
- b. 顯示感染病例人數最多的5個國家，並按遞減順序顯示
- c. 顯示台北市各區病例人數
- d. 顯示台北市最近病例的日期



305 登革熱病例統計

- 輸入說明：

讀取read.csv

- 輸出說明：

分別輸出下列四段資料：

- a. 以遞減順序顯示居住縣市的病例人數
- b. 顯示感染病例人數最多的5個國家，並按遞減順序顯示
- c. 顯示台北市各區病例人數
- d. 顯示台北市最近病例的日期

305 登革熱病例統計

- 範例輸入: 無
- 範例輸出:

居住縣市

台北市	76
新北市	60
台中市	43
桃園市	43
高雄市	42
台南市	20
彰化縣	14
新竹市	8
屏東縣	7
南投縣	6
新竹縣	5
苗栗縣	5
嘉義縣	4
雲林縣	3
宜蘭縣	2
基隆市	2
台東縣	2
澎湖縣	2
嘉義市	1

Name: 居住縣市, dtype: int64

感染國家

越南	94
菲律賓	50
泰國	44
馬來西亞	33
緬甸	29

Name: 感染國家, dtype: int64

居住鄉鎮

中山區	8
中正區	10
信義區	7
內湖區	10
北投區	1
南港區	7
士林區	7
大同區	5
大安區	10
文山區	2
松山區	3
萬華區	6

Name: 居住鄉鎮, dtype: int64

發病日: 2018/06/25



401 資料:折線圖

- 設計說明:

利用程式內提供的數據,依下列要求以matplotlib輸出chart.png圖檔:

- 設定線條寬度設定為1, 線條色彩與樣式各別為藍色虛點線與紅色虛點線
- 設定軸刻度調整至(0,0)開始至(8,70)
- 標題設定
 - 圖表標題:Figure 字體大小:24
 - X軸標題:x-Value, 字體大小:16
 - Y軸標題:y-Value, 字體大小:16

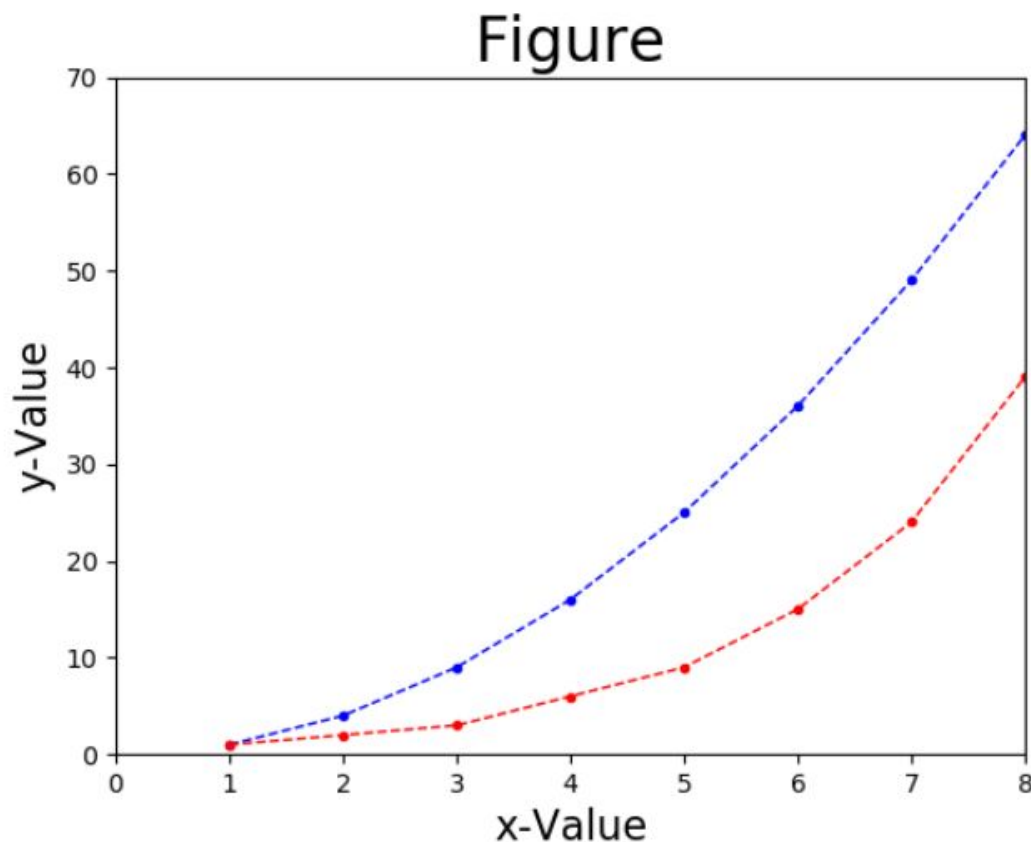


401 資料:折線圖

- 輸入說明:
無
- 輸出說明:
輸出chart.png圖檔

401 資料:折線圖

- 範例輸入:無
- 範例輸出:



注意:

matplotlib套件的版本, 會造成輸出的圖檔有差異, 但不影響評分的準確性。
Code Judger平台會將您的程式, 於伺服器中運行輸出圖檔進行評分。



402 市場成交行情:折線圖

- 設計說明：

請讀取果菜市場香蕉成交行情read.csv資料，主要有兩個欄位：成交日期、成交平均價。再以matplotlib輸出折線圖chart.png，輸出圖表的參數如下：

- 顯示圖例(legend): banana
- 圖表標題: Market Average Price
- 以成交日期為X軸, X軸名稱: date
- 以成交平均價為Y軸, Y軸名稱: NT\$
- Y軸下限15、上限25

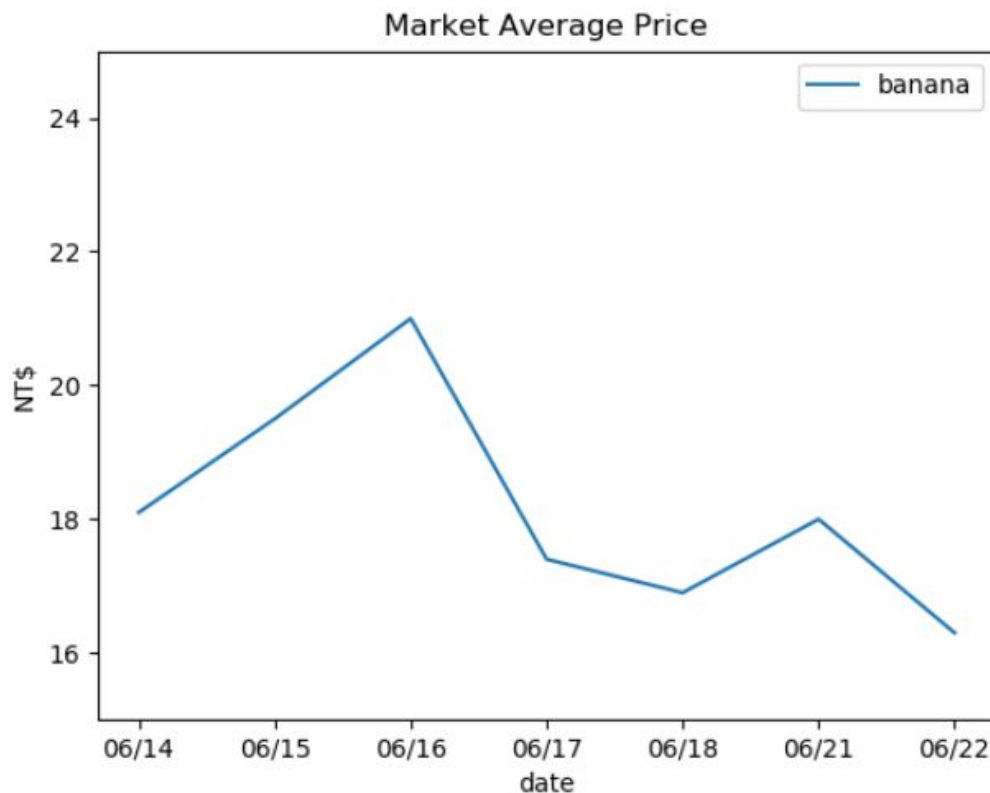


402 市場成交行情:折線圖

- 輸入說明:
讀取read.csv的內容
- 輸出說明:
輸出chart.png圖檔

402 市場成交行情：折線圖

- 範例輸入：無
- 範例輸出：



注意：

matplotlib套件的版本，會造成輸出的圖檔有差異，但不影響評分的準確性。

Code Judger平台會將您的程式，於伺服器中運行輸出圖檔進行評分。



403 月份統計:長條圖與圓餅圖

- 設計說明:

依下列要求以matplotlib輸出四個月份 (labels = 'Jun', 'Jul', 'Aug', 'Sep') 的統計圖形chart.png, 輸出圖表的參數如下:

- 完成左右兩個圖, 左圖為長條圖 (bar), 右圖為圓餅圖 (pie)
- 長條圖以labels為X軸, sizes為Y軸, 各長條顏色為藍色 (blue)
- 圓餅圖以labels為圖標, sizes為各項所占百分比
- 圓餅圖colors為各項顏色, 長寬比為1:1, 並突顯「Aug」
- 圓餅圖顯示各項百分比到小數點第1位

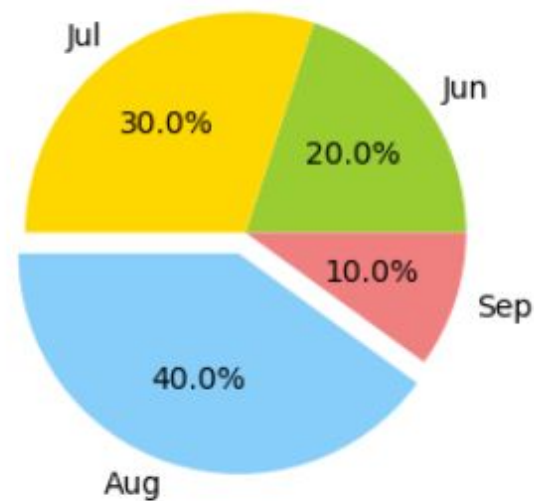
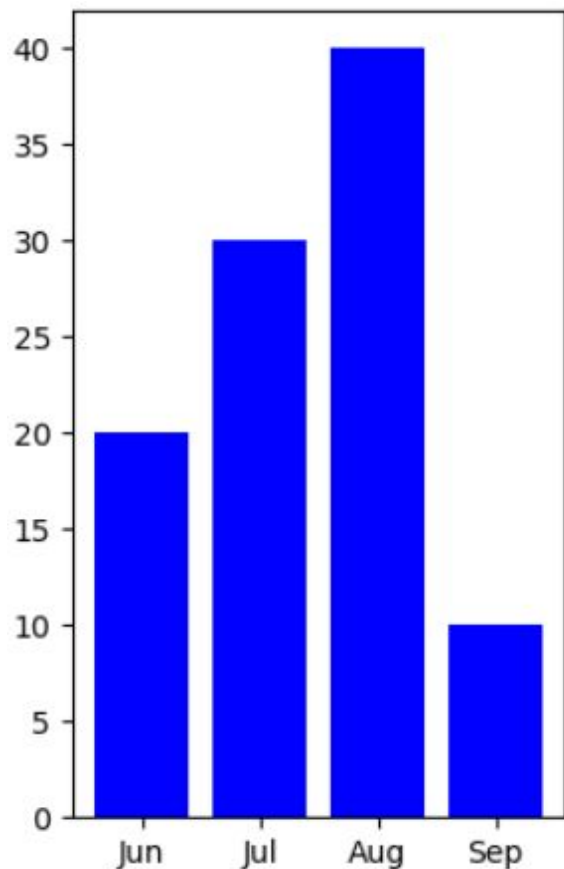


403 月份統計:長條圖與圓餅圖

- 輸入說明:
無
- 輸出說明:
輸出chart.png圖檔

403 月份統計:長條圖與圓餅圖

- 範例輸入:無
- 範例輸出:





404 成績統計:長條圖

- 設計說明:

請讀取read.csv中的資料,再以matplotlib輸出長條圖chart.png,輸出圖表的參數如下:

- 圖表標題:Score ranges count
- X軸名稱:Range
- Y軸名稱:Quantity
- 標題字型大小:20
- X軸和Y軸字型大小:14
- 長條寬度:2
- X軸刻度:0~19, 20~39, 40~59, 60~79, 80~100
- Y軸刻度:0到25, 間隔5

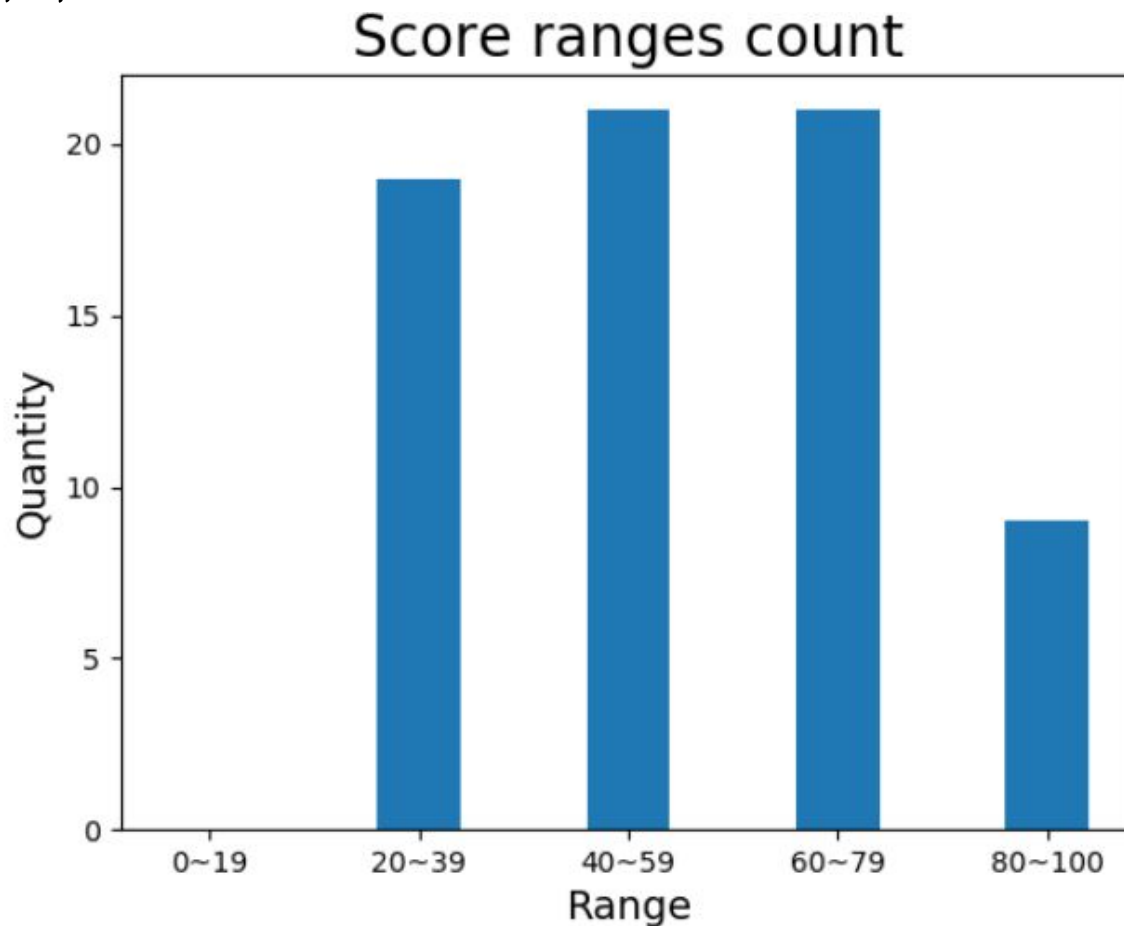


404 成績統計:長條圖

- 輸入說明:
讀取read.csv的內容
- 輸出說明:
輸出chart.png圖檔

404 成績統計:長條圖

- 範例輸入:無
- 範例輸出:





405 樣本:直方圖與散佈圖

- 設計說明:

依下列要求以matplotlib輸出chart.png圖檔:

- 利用程式內提供的第一組sample1與第二組sample2數據輸出左右兩個圖, 左圖為直方圖(histogram), 右圖為散佈圖(scatter)。
- 直方圖疊合兩個直方圖, 兩圖均請在-3~+3間均勻間隔分為100格, 透明度(alpha)均為0.5。在左上角放置圖例, sample1的標記為samples 1, sample2的標記為samples 2。
- 散佈圖以 sample1 作為X軸資料、sample2作為Y軸資料, 透明度設為 0.2。



405 樣本:直方圖與散佈圖

- 輸入說明:
無
- 輸出說明:
輸出chart.png圖檔

405 樣本:直方圖與散佈圖

- 範例輸入:無
- 範例輸出:

