

商管程式設計 (106-2)

作業八

作業設計：孔令傑

國立臺灣大學資訊管理學系

繳交作業時，請至 PDOGS (<http://pdogs.ntu.im/judge/>) 為第一題上傳一個 PDF 檔、為第二題做同儕互評，再為第三題與第四題各上傳一份 Python 3.6 原始碼 (以複製貼上原始碼的方式上傳)。第四題是 bonus 加分題。每位學生都要上傳自己寫的解答。不接受紙本繳交；不接受遲交。請以英文或中文作答。

這份作業的截止時間是 **2018 年 5 月 28 日凌晨一點**。在你開始前，請閱讀課本的第十四章第一至四節以及第十五至十七章¹。為這份作業設計測試資料並且提供解答的助教是鄭允頌。

第一題

(20 分；每小題 5 分) 針對以下題目，請寫 python 程式去針對第三題給定的「Bike.csv」畫出題目要求的圖。你可以不使用 `matplotlib`，但你一定要寫 python。請貼上你的 python 程式 (包含引入函式庫、讀入 CSV 檔、計算出必要的值、繪圖等所有程式碼) 和畫出來的圖到 PDF 檔中。針對每一張圖，都請適當地美化它 (例如調整兩軸的範圍、加上兩軸的標題、加上圖例等)。

- (5 分) 針對上班日 (`workingday` 為 1) 與非上班日 (`workingday` 為 0) 畫兩個長條圖，分別呈現該種上班狀態下的 `casual` 和 `registered` 的日平均值。更精確地說，第一個長條圖上應該有兩個長條，其中第一條是兩年間所有上班日的 `casual` 值的平均，第二條是兩年間所有非上班日的 `casual` 值的平均；第二個長條圖上應該有兩個長條，其中第一條是兩年間所有上班日的 `registered` 值的平均，第二條是兩年間所有非上班日的 `registered` 值的平均。請簡要地寫下你從這兩張圖中得到的資訊。
- (5 分) 針對 `casual` 和 `registered` 各畫一張直方圖，去呈現兩年間該種租賃的日租賃數量的分佈方式。針對 `casual`，請把組端點設成 0、500、1000、1500 直到 3500；針對 `register`，請把組端點設成 0、1000、2000、3000 直到 7000。你可以自行決定恰好等於組端點的值要歸屬到哪一組。請簡要地寫下你從這兩張圖中得到的資訊。
- (10 分) 畫一個折線圖，去呈現逐月的 `casual`、`registered` 與 `cnt` 的總和變化趨勢。更精確地說，圖上應該有三條折線，每條折線上有 24 個點，每個點代表該月在 `casual`、`registered` 或 `cnt` 的總和。請簡要地寫下你從這張圖中得到的資訊。

第二題

(20 分) 請在 PDOGS 上批改你被隨機分配到的作業七第一題的程式碼，根據它在正確性以外的部份給它 1 至 5 分的評分，並且說明你給分的依據。建議在評分時參考以下六個面向。在前五個面向上，一個面向上做得好就得一分，還不錯則半分，不好則零分；在第六個面向上則在有必要時扣分。六個面向的分數合計後無條件進入即為你最後給的總分。

¹課本是 A. Downey 所著的 *Think Python 2*，在 <http://greenteapress.com/wp/think-python-2e/> 可以下載。

- 可讀性：變數與函數名稱是否具有合適的資訊量？程式碼排版是否良好且具有前後一致性？是否有合適的註解？關於註解，當然不需要每一行都有註解，但若你發現在某一大段落裡都沒有註解，或某個你感覺很不易看懂的部份沒有註解，你可以指出來；不要直接說「註解太少」但沒有說是哪邊缺乏註解。
- 模組化程度：是否有宣告合適的函數？是否有避免將非常類似的程式片段寫複數次而非寫成函數？是否有避免一個函數做非常多事情？函數間是否有合適的 decoupling？直接閱讀程式是否能很快地理解程式在大方向上的運算邏輯？
- 效率：程式運算是否有合理的運算效率？當然我們不要求每個同學都寫出超級有效率的精妙演算法，但至少一個程式不應該進行過多不必要的運算，也不應該耗用過多不必要的記憶體空間。如果你看不出這個程式的效率有明顯的問題，我們建議你直接給一分。
- 擴充性：當要解的問題變得更複雜的時候，我們能不能簡單地修改這個程式以解決新的問題，而不是寧可砍掉重練？這個議題當然也很主觀，所以如果你不能明確地指出在怎樣的新問題上，這個程式會有擴充性問題，我們建議你直接給一分；如果你不能指出很嚴重的問題，我們建議你至少給半分。但對批改者來說，這個關於擴充性的思考其實是很好的訓練。試試看吧！
- 其他：如果有任何其他令你想扣分的理由，請明確地寫出來並且在這個面向上扣分；沒有的話就給一分。
- 題目規範：你應該檢查那份程式碼有沒有違反題目的規範，如果有（例如題目說不可以用上課沒教過的東西，但他用了，或者題目說一定要用指標和動態記憶體配置，但他沒用），就扣他三分。當然，請明確地指出他哪邊違反了題目的規範。

本題其中 10 分取決於檢視你的程式碼的同學給你的分數總和（必要時助教會出來主持公道，請不用緊張），另外 10 分取決於你對同學的程式碼的評語和評分的合理性和建設性（原則上除非被申訴，且助教檢視後發現你確實評得很不公，否則只要有評就會得到 10 分）

第三題

（60 分）在隨本題附上的「Bike.csv」檔中，我們有從 2011/1/1 到 2012/12/31 在美國某一城市的公共自行車租賃系統的每日租賃記錄，檔案中有 731 列，代表時間範圍內的 731 天，部份資料如表 1 所示。

date	holiday	workingday	weathersit	temp	atemp	humidity	windspeed	casual	registered	cnt
2011/1/1	0	0	2	14.11	18.18	80.58	5.94	331	654	985
2011/1/2	0	0	2	14.9	17.69	69.61	9.20	131	670	801
2011/1/3	0	1	1	8.05	9.47	43.73	9.19	120	1229	1349
2012/12/29	0	0	2	10.39	12.12	75.29	4.60	159	1182	1341
2012/12/30	0	0	1	10.49	11.59	48.33	12.98	364	1432	1796
2012/12/31	0	1	2	8.85	11.17	57.75	5.73	439	2290	2729

表 1: 逐日租賃記錄

各欄位項目解釋如下：

- （日期資訊）date：為一個 YYYY/MM/DD 字串，但月或日若為個位數，則只呈現一位數字，例如 2011/1/1 不會儲存成 2011/01/01。

- (工作相關資訊) holiday、workingday：holiday 為 1 表示那天是除了週末以外的國慶假期，否則為 0；workingday 為 1 表示那天是要上班的工作日（既不是週末也不是其他假日），否則為 0。
- (天氣相關資訊) 此類別中記錄了五個屬性。weathersit (weather situation) 為 1 表示陽光或部分多雲，2 表示陰霾和多雲，3 表示輕雪或小雨。temp (temperature) 與 atemp (apparent temperature) 表示當天的平均溫度和平均體感溫度（攝氏）。humidity 表示當日平均濕度（%）。windspeed (wind speed) 表示風速的日平均值（以節為單位，一節約 1.852 公里/小時）。
- (租賃數據) casual、registered、cnt：casual 代表來自未註冊的使用者的當日總租賃數量、registered 是來自註冊會員的當日總租賃數量、cnt 是兩個數字的和。

給定這個檔案，以及一些篩選條件，我們想要彙總數天資料以取得一個日期區間內的租賃總額。篩選條件包含起始年月、終止年月、星期幾與天氣代碼。請先取出所有介於起始年月與終止年月（包含這兩個月）之間的每一天，從中取出該日期符合指定之星期幾且該日天氣符合天氣代碼的日子，最後將這些天的 casual 與 registered 分別加總並輸出。

舉例來說，如果給定起始年月為 2011 年 1 月、終止年月為 2011 年 2 月，指定週一與晴天，則符合條件的共有 1/3、1/10、1/24、2/7、2/14，他們的 casual 與 register 加總各為 575 與 7136。

輸入輸出格式

系統會提供一共 20 組測試資料，每組測試資料裝在一個檔案裡。在每個檔案中會有兩行，第一行包含一個字串，代表需讀取之檔案的絕對路徑；第二行包含兩個字串，依序代表指定的開始年月與結束年月（所以前者不晚於後者）年月的格式為 yyyy/mm，且月份必定以兩位數字呈現，例如 2012 年 8 月表示為 2012/08。兩個時間字串之間被一個空白隔開。第三行包含兩個整數，分別是星期幾的代碼以及天氣代碼，其中星期幾的代碼 1 表示週一、2 表示週二，依此類推，7 表示週日；天氣代碼的 1 到 3 則和 weathersit 欄位中記錄的 1 到 3 有同樣意義。我們只會給你隨本題附上的這個「Bike.csv」檔案。

讀入資料後，請按照题目的規定篩選出在該日期區段內星期幾和天氣符合篩選條件的日期，並印出這些日期的 casual 與 registered 欄位的和。這兩個數字之間用一個空白鍵隔開。舉例來說，如果輸入是

```
C:/thisIsAnExamplePath/Bike.csv
2011/01 2011/02
1 1
```

則輸出應該是

```
575 7136
```

如果輸入是

```
C:/thisIsAnExamplePath/Bike.csv
2011/01 2011/01
1 2
```

則輸出應該是

```
159 2342
```

你上傳的原始碼裡應該包含什麼

你的.py 原始碼檔案裡面應該包含讀取測試資料、做運算，以及輸出答案的 Python 3.6 程式碼。當然，你應該寫適當的註解。針對這個題目，你**不可以**使用上課沒有教過的方法，但上課介紹過的函式庫中所有的功能都可以用。

評分原則

- 這一題的其中 40 分會根據程式運算的正確性給分。PDOGS 會直譯並執行你的程式、輸入測試資料，並檢查輸出的答案的正確性。一筆測試資料佔 2 分。
- 這一題的其中 20 分會在作業九中被評定。屆時我們會讓同學們互相檢視彼此的本題程式碼，並且就可讀性、易維護性、模組化程度、排版等面向寫評語和給評分（當然一切都是匿名的）。該任務在本題中會佔 20 分，其中 10 分取決於檢視你的程式碼的同學給你的分數（必要時助教會出來主持公道，請不用緊張），另外 10 分取決於你對同學的程式碼的評語和評分的合理性和建設性。若你在本次作業中完全沒有寫這一題，那屆時自然沒有人能檢視你的程式碼，你也就得要損失這 10 分了。

第四題 (bonus)

(20 分) 承上題，我們繼續對「Bike.csv」做運算。現在在給定起始月份與終止月份後，我們想要產生一張以天氣代碼和上班日、非上班日為維度的樞紐分析表，一共有三列兩欄，每一列是一個天氣代碼（依序由 1 到 3），第一欄是上班日，第二欄是非上班日，表中的每一個格子則儲存在指定的起始和終止月份（包含這兩個月）中，所有上班型態符合該欄且天氣符合該列天氣的日子的 casual 或 registered 的和。題目會指定要輸出 casual 還是 registered 的和。舉例來說，如果要針對 2012/1 到 2012/10 間的 casual 做分析，我們會得到

Weather label	Working situation	
	Working	Non-working
1	122645	123853
2	45771	44349
3	317	1576

系統會提供一共 10 組測試資料，每組測試資料裝在一個檔案裡。在每個檔案中會有三行，第一行包含一個字串，代表需讀取之檔案的絕對路徑；第二行包含兩個字串，依序代表指定的開始年月與結束年月（所以前者不晚於後者）年月的格式為 yyyy/mm，且月份必定以兩位數字呈現，例如 2012 年 8 月表示為 2012/08。兩個時間字串之間被一個空白隔開。第三行包含一個整數，若為 1 則表示針對 casual 做分析，若為 2 則表示針對 registered 做分析。我們只會給你隨本題附上的這個「Bike.csv」檔案。

讀入資料後，請按照題目的規定印出三行，第 i 行代表針對天氣代碼 i 的分析結果，每一行有兩個數字，依序是該時間區段中該天氣下的工作日與非工作日的 casual 或 registered 總和，中間用一個空白字元隔開。每一行（包含第三行）的最後面不應該有空白字元，應該要有一個換行字元。

舉例來說，如果輸入是

```
C:/thisIsAnExamplePath/midterm2.csv  
2012/01 2012/10  
1
```

則輸出應該是

```
122645 123853  
45771 44349  
317 1576
```

針對這個題目，你**可以**使用任何方法。這一題的 20 分都根據程式運算的正確性給分，一筆測試資料佔 2 分。