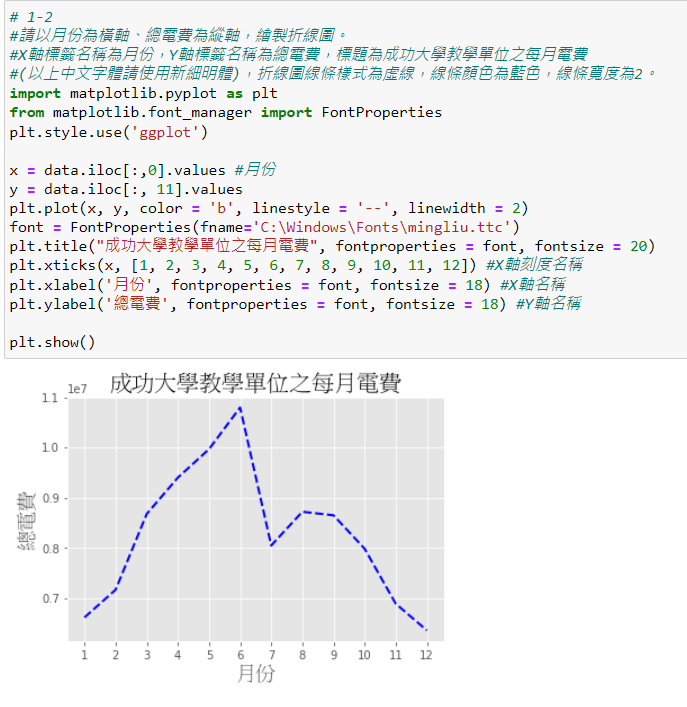
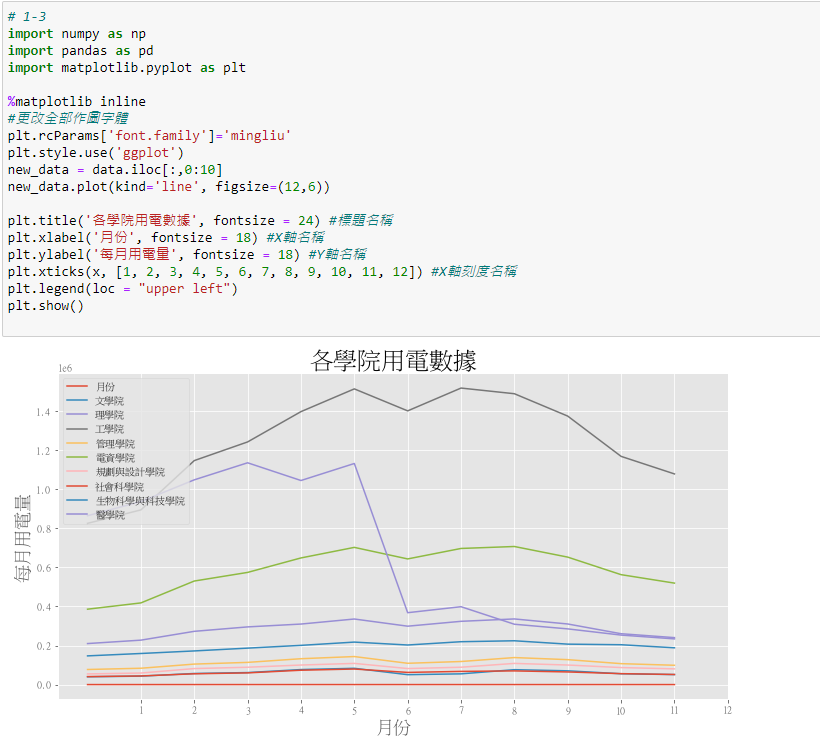
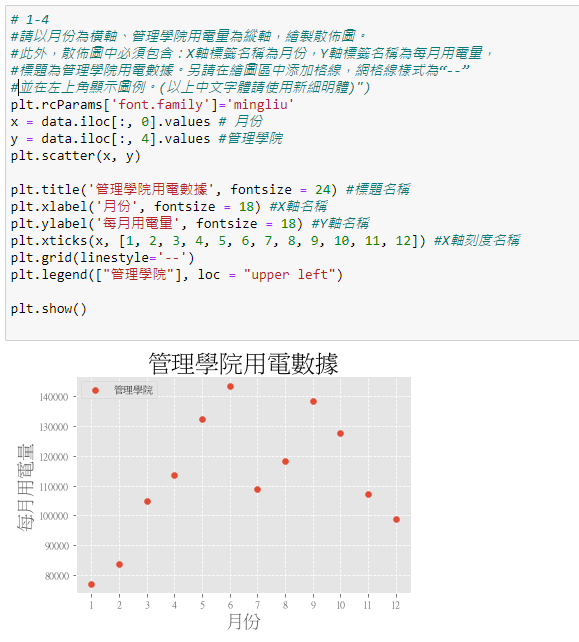
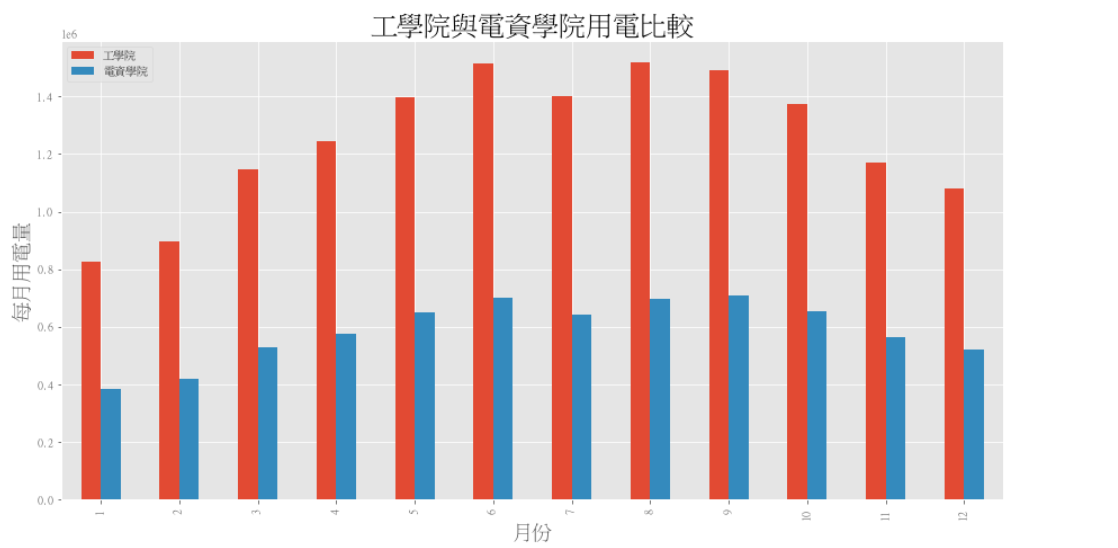
大數據分析與資料探勘-第二次作業

繳交日期: 2022/3/29

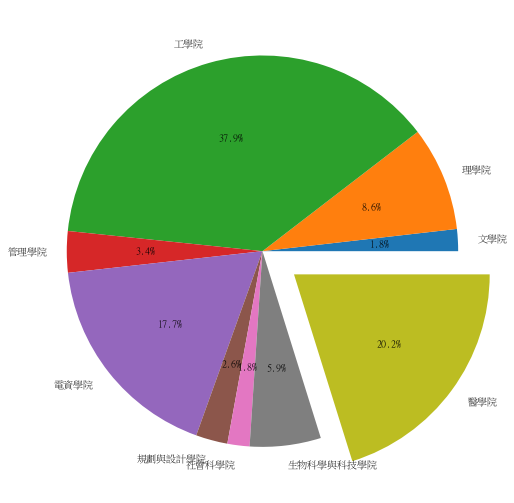
使用Python程式進行分析並回答以下所有問題。請直接在word檔中每個問題的下面，貼上python程式與執行結果，最後將檔案上傳到MOODLE，包含此word檔及python程式檔。兩個檔案，麻煩請壓縮成一個檔在上傳，檔名請命名為Homework2\_姓名.zip，另程式碼中請詳細標明對應的題數，若看不懂的話，一律不給分。上傳期限為3/29晚上11點59分前，超過繳交期限就無法上傳。

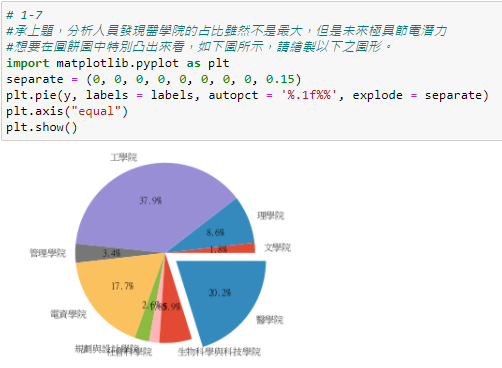
1. 請匯入NCKU\_ele\_data資料檔至DataFrame，並加以顯示。
2. 請以月份為橫軸、總電費為縱軸，繪製折線圖。此外，在此折線圖中必須包含：X軸標籤名稱為月份，Y軸標籤名稱為總電費，標題為成功大學教學單位之每月電費(以上中文字體請使用新細明體)，折線圖線條樣式為虛線，線條顏色為藍色，線條寬度為2。
3. 請以月份為橫軸、各學院用電量為縱軸，以折線圖繪製九大學院的每月用電量(九條折線圖)。此外，折線圖中必須包含：X軸標籤名稱為月份，Y軸標籤名稱為每月用電量，標題為各學院用電數據。另請在左上角顯示圖例(標示那一條線為哪個學院的用電量)。(以上中文字體請使用新細明體)
4. 請以月份為橫軸、管理學院用電量為縱軸，繪製散佈圖。此外，散佈圖中必須包含：X軸標籤名稱為月份，Y軸標籤名稱為每月用電量，標題為管理學院用電數據。另請在繪圖區中添加格線，網格線樣式為“--”，並在左上角顯示圖例。(以上中文字體請使用新細明體)
5. 一張含有 文字 的圖片

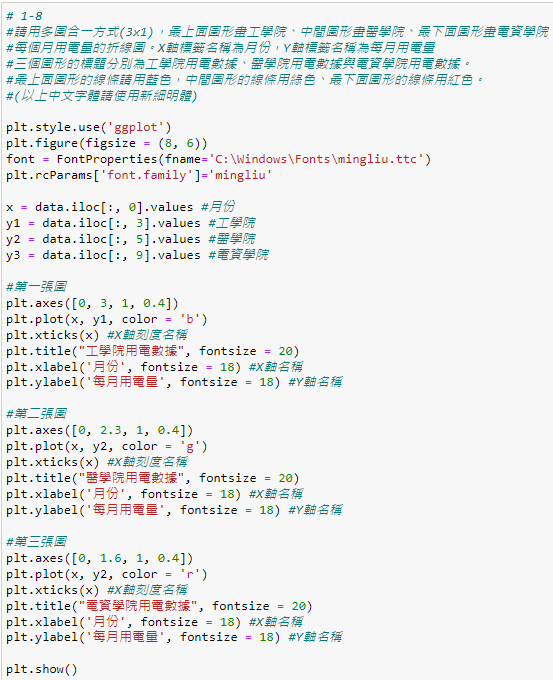
   自動產生的描述請以月份為橫軸、每月用電量為縱軸，以長條圖繪製工學院與電資學院這兩種學院的每月用電量比較(亦即繪製兩條長條圖)。此外，長條圖中必須包含：X軸標籤名稱為月份，Y軸標籤名稱為每月用電量，標題為工學院與電資學院用電比較，並在左上角顯示圖例。(以上中文字體請使用新細明體)
6. 請把9大學院12個月的用電量加總為一整年的用電量，然後計算每個學院年用電量占年總用電量的比例，繪製成圓餅圖，請把每個學院名稱及占比顯示在圖上。(以上中文字體請使用新細明體、占比請顯示小數點1位)

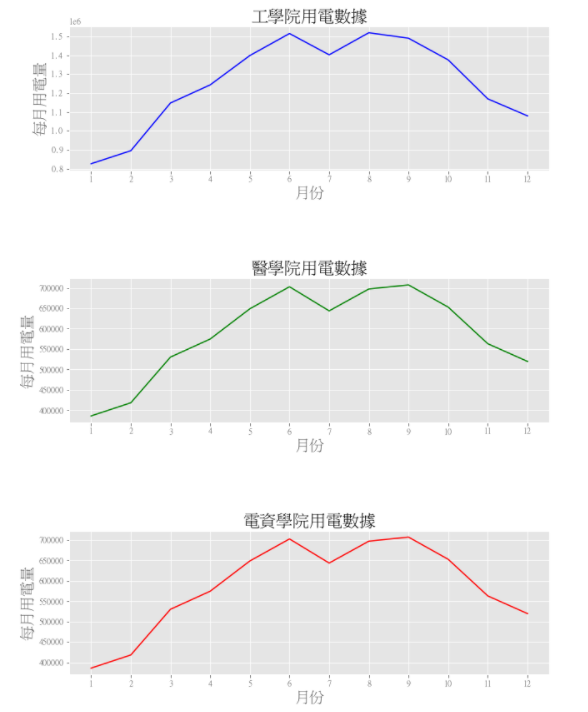


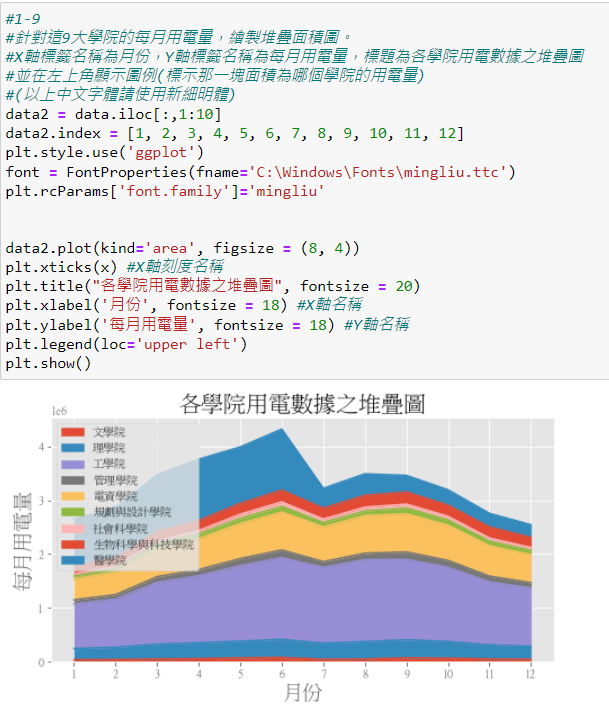
1. 承上題，分析人員發現醫學院的占比雖然不是最大，但是未來極具節電潛力，想要在圓餅圖中特別凸出來看，如下圖所示，請繪製以下之圖形。





1. 請用多圖合一方式(3x1)，最上面圖形畫工學院、中間圖形畫醫學院、最下面圖形畫電資學院每個月用電量的折線圖。此外，圖形中必須包含：X軸標籤名稱為月份，Y軸標籤名稱為每月用電量，三個圖形的標題分別為工學院用電數據、醫學院用電數據與電資學院用電數據。最上面圖形的線條請用藍色，中間圖形的線條用綠色、最下面圖形的線條用紅色。(以上中文字體請使用新細明體)



1. 針對這9大學院的每月用電量，繪製堆疊面積圖。此外，堆疊圖中必須包含：X軸標籤名稱為月份，Y軸標籤名稱為每月用電量，標題為各學院用電數據之堆疊圖，並在左上角顯示圖例(標示那一塊面積為哪個學院的用電量)。(以上中文字體請使用新細明體)

Note：繪圖問題中有多項要求，請一一滿足，若僅完成一部分，則僅會得到部分分數。

2.

1. 一張含有 桌 的圖片

   自動產生的描述一張含有 文字 的圖片

   自動產生的描述請匯入health資料檔至DataFrame，並加以顯示。
2. 一張含有 文字 的圖片

   自動產生的描述在此DataFrame最後增加一個新的column，計算「每日行走步數」除上「攝取卡路里」，將此新的column名稱命名為「每單位卡路里之行走步數」，並加以顯示。

一張含有 桌 的圖片

自動產生的描述

1. 承上題，依據「每單位卡路里之行走步數」可判斷出每天的健康指數，若值小於等於2.5則顯示「低」，若值大於2.5、小於等於5則顯示「中」，若值大於5則顯示「高」，請在上題的DataFrame最後再增加一個新的column，顯示這些健康指數，並將此新的column名稱命名為「健康指數」。

一張含有 文字 的圖片

自動產生的描述一張含有 桌 的圖片

自動產生的描述

1. 一張含有 文字 的圖片

   自動產生的描述承上題，由於這些健康指數為數值型資料，使用One-hot Encoding方法將其擴展為三個columns (亦即對應於該列的column為1，其餘columns為0)，將這三個新columns增加到上題DataFrame的最後，column名稱分別為「健康指數\_高」，「健康指數\_中」和「健康指數\_低」。

一張含有 桌 的圖片

自動產生的描述