

資料庫管理（114-1）

作業一

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 NTU COOL 下載本作業題目的地方上傳一個 PDF 檔。在生成這個 PDF 檔時，可以用打字的也可以用手寫的，但不管怎樣，請務必注意繳交的文件的专业程度（通常透過排版、文字圖片表格方程式的清晰程度、用字遣詞等面向呈現），如果專業程度不夠，也會被酌量扣分。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。可以用中文或英文作答。這份作業的截止時間是 **9 月 10 日早上 08:00:00**，遲交在 12 小時內者扣 10 分、在 12 到 24 小時內者扣 20 分、超過 24 小時的這份作業將得不到分數。

相關規定與提醒

1. 關於上網查詢與 AI 工具：

任何一份作業都可以被用任何方式完成，包括上網搜尋和使用各種 AI 工具。如果你想用，請留意以下幾件事。首先，抄襲還是不被允許的，如果我們發現抄襲（包括抄襲網路上的答案，或是抄襲同學的答案），都還是會給予嚴厲的懲罰（視情節輕重而定，通常是該份作業算零分，或者不予通過這門課）。只要沒有抄襲，那我們就只根據你交上來的答案的品質給分，不論你是自己想出來的，還是有利用 AI 工具。如果某甲善用了 AI 工具後寫得很好，某乙自己努力寫但寫得不好，那某甲會得到比較高分。其次，如大家所知，AI 工具給的答案可能會錯，也可能不合適。使用 AI 工具是學生的自由，但確認 AI 工具的答案是否合適、是否需要調整則是學生的任務。請務必自行確認答案的正確性與合適性。最後還是提醒大家，學到多少東西都是自己的，如果一時困難用 AI 工具度過難關那是無妨，但之後建議還是花時間把東西學起來，對自己比較好。

2. 關於「專業」：

在我們這門課及許多課程中，都需要繳交各種報告。一份報告如果要達成他的效果（例如成功募資、完成溝通的任務等等），除了需要好的內容，也需要「專業」，而顯得專業通常需要「長得好看」以及「看起來用心」，這在沒有標準答案的任務上更是如此。有鑑於此，在這門課的作業和專案，我們都會要求報告的格式和美觀，並且納入評分標準。為此，我們提供報告格式參考指南「DB_reportFormatGuideline.pdf」，上面列舉了一份格式良好的報告的最低標

準。在作業一我們會請助教就違反參考指南的地方標出來讓大家知道（我們理解那份指南並不是最完美的，但如果完全沒有標準，同學們容易無所適從，所以我們還是設計一份當標準），但不會扣分，只是提醒大家；從作業二起就會有部分分數是報告專業度分數。之所以要求這些不是想要找大家麻煩，而是大家離出社會也不是太遠了，確實應該要開始被要求報告的可讀性和易讀性，所以我們願意花一些時間要求大家，但不會刁難大家，也請大家理解和盡力嘗試了。

3. 關於「批改」：

如課程大綱所述，為了不要累死助教，每次作業可能只有部分題目會被批改和給予回饋，但每一題的參考解答都會在作業截止後公佈。如果有一題沒被批改，那所有有寫那一題的學生都會得到那一題的全部分數，但沒寫或期限前沒交作業的自然就不會拿到那一題的分數。最後，請注意是「可能」，換言之也有可能是所有題目都被批改。

4. 關於「作業一」：

設計作業一事實上就只有一個目的：讓大家思考一下自己是否有「程式設計」、「資料結構」和「作業系統」的基礎知識，如果沒有，要嘛不要修這門課，不然就是在需要的時候要花額外的時間補充，或者放棄聽懂部份主題（和分數）。大家當然可以用 AI 工具寫這一份作業，只要你看得懂 AI 在說什麼、有辦法判斷它說的是對是錯，那應該就算通過作業一的挑戰。但從作業二起，會有相對更多題目是 AI 工具比較無法回答的，這部份就請大家自行斟酌了。

題目

1. (45 分；每小題 15 分) 在「Bike.xlsx」中，我們有從 2011/1/1 到 2012/12/31 在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。各欄位項目解釋如下：

- (工作相關資訊) holiday、weekday、workday：holiday 為 1 表示那天是除了週末以外的國慶假期，否則為 0；weekday 為 0 表示當天是星期日，為 1 表示當天是星期一，依此類推；workday 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。

請寫程式分析這批資料（可以直接讀取 XLSX 檔，也可以把 XLSX 檔轉成 CSV 或 TXT 檔後再讀取；可以使用任何你喜歡的程式語言）並回答以下問題。除了寫下你的答案，也請貼上你的程式碼。

- 有一個假說是「會員需求在上班日比非上班日高、非會員需求在非上班日比上班日高」。請用最簡單的敘述統計（平均數、標準差、樣本數那一類的）檢驗這個假說，並說明你是否支持這個假說。你不需要做任何推論統計中的假設檢定。
- 有一個假說是「不論哪一個月，會員總需求總是比非會員總需求高」。請用最簡單的敘述統計（平均數、標準差、樣本數那一類的）檢驗這個假說，並說明你是否支持這個假說。你不需要做任何推論統計中的假設檢定。
- 有一個假說是「第二年的需求變異程度高於第一年」。請用最簡單的敘述統計（平均數、標準差、樣本數那一類的）檢驗這個假說，並說明你是否支持這個假說。你不需要做任何推論統計中的假設檢定。

2.（35 分）請簡要地回答以下問題：

- （10 分）針對複習影片「An application of pointers: sorting C strings alphabetically」中的字串排序，請用自己的話說明為什麼「交換指標」的排序法會比「交換字串」更有效率。請將你的答案限制在 300 字內。
- （10 分）請把以下數字依序放入一個 max heap，並把最終結果畫出來：1、5、16、18、13、7、19、2、27、13、35、4、20。
- （15 分）請自行學習資料結構中的「Left-child-right-sibling」表示法，然後用自己的話說明是否可以用 binary tree 表達任意一棵 non-binary tree（所謂可以的意思就是，針對一個給定的 non-binary tree，你畫另外一棵 binary tree，然後只要看著你畫的 binary tree，就可以回推得到原本的 non-binary tree 上的所有資訊）。請將你的答案限制在 300 字內。

3.（20 分；一題 5 分）請簡要地回答以下問題，每題最多 200 字：

- 「作業系統」這門課通常涵蓋哪些主題？

- (b) 請用自己的話說明什麼是「哲學家用餐問題」、該問題跟作業系統中的什麼議題有關，以及任意一個常見的解法。
- (c) 磁碟管理中的連續分配（contiguous allocation）原則是什麼，又有什麼優缺點？
- (d) 為什麼作業系統會有「虛擬記憶體」這項技術？其主要用途是什麼？