大數據分析與資料探勘-第一次作業

繳交日期: 2023/3/22

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、python程式檔及匯出excel檔案(供佐證用)。上述3個檔案，麻煩請壓縮成1個檔在上傳，檔名請命名為Homework1\_姓名.zip (如：Homework1\_黃韻勳.zip)，另程式碼中請標明對應的題數(如：#題數)，若看不懂的話，一律不給分；若檔案有缺少，亦會加以扣分。上傳期限為3/22晚上11點59分前，超過繳交期限就無法上傳。

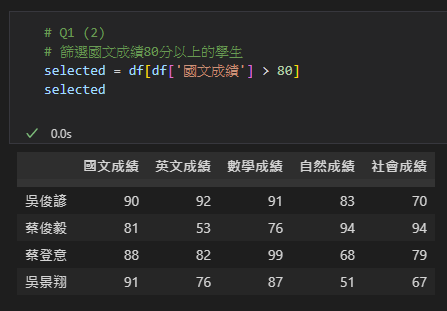
1. 請用手動建立一個DataFrame如下表，並加以顯示。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 國文成績 | 英文成績 | 數學成績 | 自然成績 | 社會成績 |
| 李國發 | 65 | 72 | 78 | 93 | 56 |
| 吳俊諺 | 90 | 92 | 91 | 83 | 70 |
| 蔡俊毅 | 81 | 53 | 76 | 94 | 94 |
| 姚鈺迪 | 79 | 85 | 47 | 88 | 80 |
| 袁劭彥 | 70 | 61 | 32 | 70 | 88 |
| 蔡登意 | 88 | 82 | 99 | 68 | 79 |
| 吳景翔 | 91 | 76 | 87 | 51 | 67 |
| 邱孝信 | 55 | 64 | 62 | 78 | 52 |
| 陳明輝 | 40 | 46 | 55 | 60 | 71 |

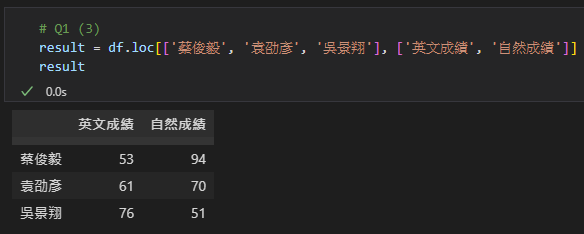
一張含有 文字 的圖片

自動產生的描述

1. 請挑出國文80分以上的學生，列出這些學生全部科目的成績

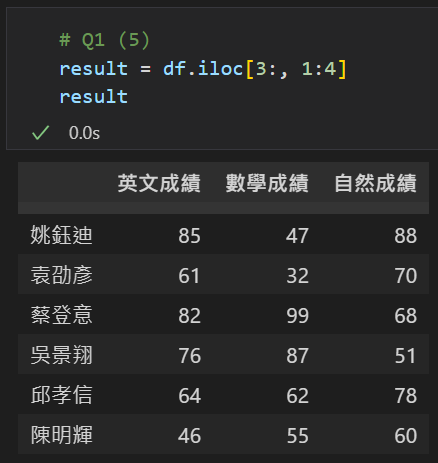


1. 請以.loc[]語法列出蔡俊毅、袁劭彥及吳景翔的英文與自然成績



1. 一張含有 文字 的圖片

   自動產生的描述請以.iloc[]語法列出吳俊諺、陳明輝的國文與社會成績
2. 請以.iloc[]語法列出姚鈺迪到最後一位學生的英文至自然的成績



1. 請以數學成績做遞增排序後，列出所有學生的全部科目成績。



1. 請計算每個科目成績的幾何平均數

一張含有 文字 的圖片

自動產生的描述

1. 現有這幾位學生的體育成績如下表，請將體育成績與第一小題的五科成績(國、英、數、自、社)加以合併為新的DataFrame (體育成績列在最後一欄)

|  |  |
| --- | --- |
|  | 體育成績 |
| 李國發 | 65 |
| 吳俊諺 | 75 |
| 蔡俊毅 | 71 |
| 姚鈺迪 | 69 |
| 袁劭彥 | 70 |
| 蔡登意 | 98 |
| 吳景翔 | 81 |
| 邱孝信 | 59 |
| 陳明輝 | 70 |

一張含有 文字 的圖片

自動產生的描述

1. 承上題，請將所有學生之全部科目成績(共有六科，包含體育)匯出為Excel檔案(.xlsx)

一張含有 文字 的圖片

自動產生的描述

2.

1. 鐵達尼號(titanic)為資料學習中常被使用的資料庫，請匯入titanic資料檔至DataFrame，並加以顯示。其中，各欄位代表的意義，分別簡述如下：Survived (0 = No; 1 = Yes)、Pclass - Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)、Names、Sex、Age、SibSp - Number of Siblings/Spouses Aboard、Parch - Number of Parents/Children Aboard、Ticket - Ticket Number、Fare - Passenger Fare、Cabin、Embarked - Port of Embarkation (C = Cherbourg; Q = 一張含有 文字 的圖片

   自動產生的描述Queenstown; S = Southampton)
2. 一張含有 文字 的圖片

   自動產生的描述存活(Survived)的欄位中，0表示最後沒有存活，1表示最後存活下來，請計算此資料中，旅客的存活比率。
3. 請計算出每個欄位分別有多少個空缺值

一張含有 資料表 的圖片

自動產生的描述

1. 一張含有 文字 的圖片

   自動產生的描述假如空缺值超過三百個，則視為資料不足的程度非常嚴重，因此請刪除空缺值超過三百個的欄位，並重新顯示新的資料表。
2. 承上題，在新的資料表中請列出Age欄位有空缺值的資料(全部欄位)

一張含有 文字 的圖片

自動產生的描述

1. 一張含有 文字 的圖片

   自動產生的描述請將Age欄位有空缺值的資料用「平均數」填補，並秀出填補後的DataFrame
2. 一張含有 文字 的圖片

   自動產生的描述承上題，請繼續將Embarked欄位有空缺值之資料用「眾數」填補，並秀出填補Age與Embarked欄位後的DataFrame
3. 一張含有 文字 的圖片

   自動產生的描述承上題，請以填補Age與Embarked欄位後的DataFrame，計算male(男生)和female(女生)的平均年齡，並加以顯示
4. 一張含有 文字 的圖片

   自動產生的描述承上題，請將male(男生)且Age小於12的Fare欄位打八折，請顯示打折之後的Fare欄位結果
5. 一張含有 文字 的圖片

   自動產生的描述承上題，請將Age大於40的Fare欄位打九折，並四捨五入到小數第一位，請顯示打折之後的Fare欄位結果

3.

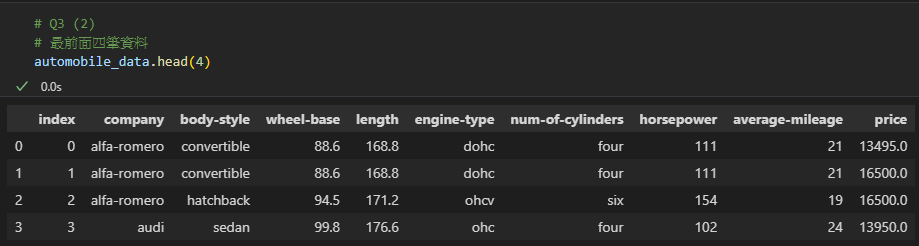
1. 請匯入Automobile\_data資料檔至DataFrame，並加以顯示

一張含有 文字, 螢幕, 螢幕擷取畫面, 屏幕、螢幕 的圖片

自動產生的描述

1. 請秀出最前面四筆資料與最後面四筆資料

一張含有 文字, 螢幕, 屏幕、螢幕, 螢幕擷取畫面 的圖片

自動產生的描述

1. 一張含有 文字 的圖片

   自動產生的描述請找出此資料中最貴的車子，並顯示出該筆資料的所有數值
2. 請找出toyota公司所製造的汽車資料，並加以顯示
3. 一張含有 文字 的圖片

   自動產生的描述一張含有 文字, 螢幕, 屏幕、螢幕, 螢幕擷取畫面 的圖片

   自動產生的描述請計算每間公司的總車輛數，並加以顯示
4. 一張含有 文字 的圖片

   自動產生的描述請顯示出每間公司所製造的車子之行駛哩數(average-mileage)的平均值，只需顯示公司名稱與平均值
5. 一張含有 文字, 螢幕, 屏幕、螢幕, 螢幕擷取畫面 的圖片

   自動產生的描述請將整個資料表以價格(price)進行遞減排序，並加以顯示
6. 請將資料表中的汽缸數量(num-of-cylinders)從文字替換成數值，如：two = 2、three = 3…..，並加以顯示整個資料表

一張含有 文字, 螢幕, 黑色, 螢幕擷取畫面 的圖片

自動產生的描述

1. 一張含有 文字, 螢幕, 螢幕擷取畫面, 黑色 的圖片

   自動產生的描述承上題，將價格(price)欄位除以轉換後的汽缸數量(num-of-cylinders)欄位，令新的欄位名稱為ratio-of-price\_cylinders，呈現在資料表的最後一欄，並加以顯示整個資料表
2. 選出company為audi且body-style為sedan的資料，命名為df\_audi，選出company為bmw且body-style為sedan的資料，命名為df\_bmw，將資料集df\_audi和df\_bmw垂直合併，並加以顯示

一張含有 文字, 螢幕, 屏幕、螢幕, 黑色 的圖片

自動產生的描述

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