

## 目標：探討作業難度

商研一 r10741050 羅奕丞

### 一、動機：

一方面是好奇大家作業的完成度，順便評估自己的學習表現，另一方面也想給老師及助教們有關作業方面的回饋，寫作業都很辛苦了，何況是設計作業。

### 二、資料篩選：

只選取 hw0 – hw7 的繳交紀錄，排除了 Midterm 以及 Quiz。

使用到的函式庫有：

1. Pandas
2. matplotlib.pyplot

### 三、評估方法

決定用 3 個 Metric 來評估作業的表現：

- 每題平均每人提交次數(mean\_submission)
- 接受率(AC\_rate)
- 每題平均每人得分率(mean\_score)

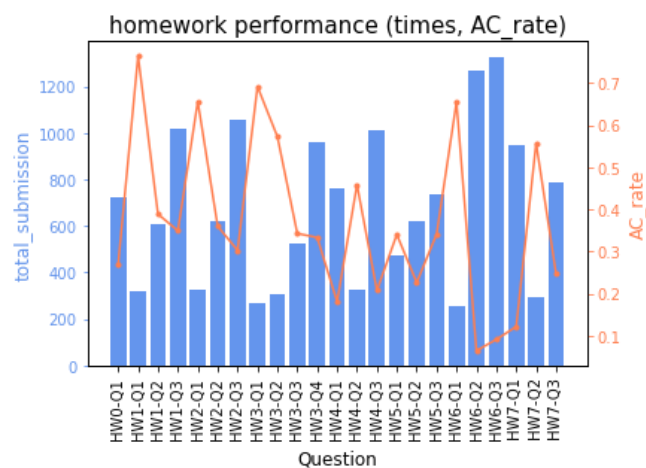
將部分程式碼、原因彙整至以下表格

Metric	計算方法	原因
mean_submission	先以 groupby 將 challenge_title 和 problem_label 組成群組，分別對 submission_id 欄位使用 count() 得到個別题目的提交次數，對 anonymous_account 欄位使用 nunique() 得到每題的提交人數，相除即得。	本來想用單題的總繳交次數，但是後來想想覺得平均每人的繳交次數應該更能夠反映個別同學的情況
AC_rate	一樣先以 groupby 將 challenge_title 和 problem_label 組成群組，接下來對 submission_result_verdict 欄位使用 value_counts() 得到相對應的次數，AC 次數除以總繳交次數即得。	AC 高可推估大家快速修正錯誤，間接反映題目難度。 更好的做法其實是追溯每位同學截止時間前的最後一筆資料，然後判斷該同學是否為 AC。
Score_rate	以 groupby 將 challenge_title 和 problem_label 和 anonymous_account 組成群組，接下來取 max() 得到每個人的最高分數，去除以該題最高分數，則得到該題的得分率。	考慮 Pdogs 測資最高分數來計算總共獲得每一小題多少比例的分數，這邊不考慮同學最後一次提交比較差的分數，畢竟曾經高分過，就以最高分數來反映真實實力吧 XD

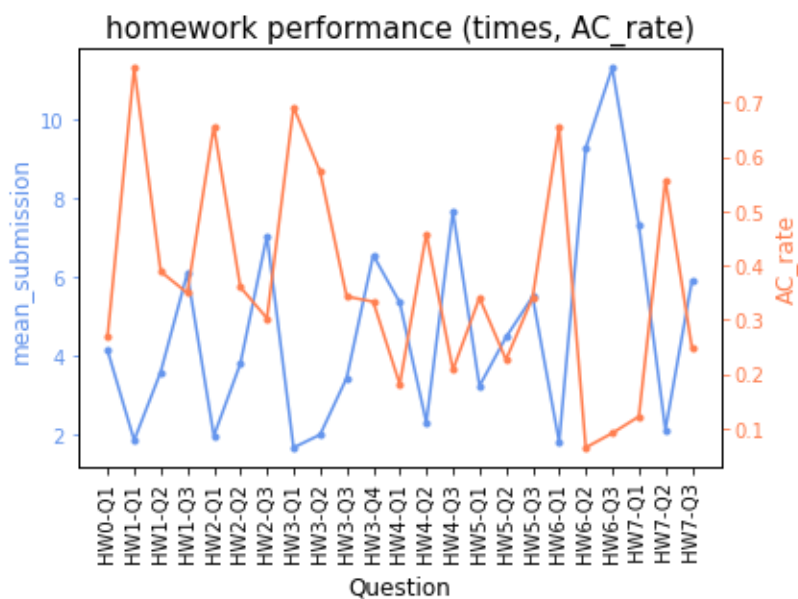
(資料處理的部分就略過)

## AC\_rate & mean\_submission

首先，我先針對 AC\_rate 畫圖，並同時考慮每一題的總提交次數(bar)，圖形如下：



結果跟預想的相差不遠，**可以看到 AC\_rate 越高的題目，通常總提交次數就越低**，反之亦然，也因此，我想到另外一個反應題目難度的 Metric，**平均每人提交次數**，可以更實際反映同學在做得作業時的痛苦情況，繪製結果如下：

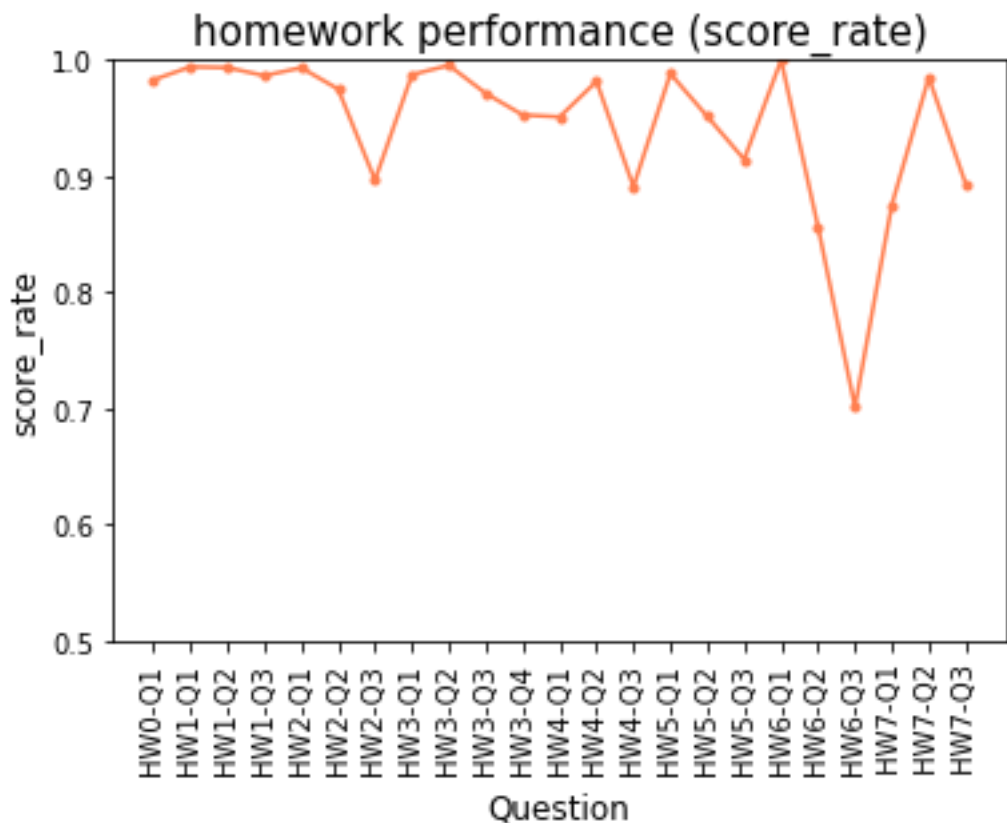


這裡我會使用折線圖的原因是更容易比較，而且我認為**平均每人提交次數**不適合用 barchart 表示，結果看來跟上一張圖的反向關係很相近，後面幾

題也差不多的情況代表停修的人可能不多 XDD 大家還是都有交作業。

## Score\_rate

接下來我們來看看得分率吧：



可以觀察到除了 Hw6 以外，大部分的得分率都有 0.9 以上。

#### 四、結論&建議

從上述的圖表可以觀察到同學明顯比較痛苦的作業題目應該是第六題，是關於**字串處理**的題目，我認為跟第一次接觸也有關係，不過我當時在寫的時候也覺得蠻難的，**還沒有搞清楚 def 就馬上要在作業中定義這麼多函式，同時又要考慮新的字串處理方法**，導致大家的 AC\_rate 較低，而且 score\_rate 的表現也較差。

至於前幾次的題目看起來相對簡單，有可能是因為之前有接觸過程式的同學課程內容相對負擔較小，或是避免一開始就打擊大家的自信心，**我要建議的部分在於，降低 Q6 字串作業的難度，或是給予 2 周的時間練習**，以我自己來說會覺得學習成效會比較好。