

人氣教授養成指南 — 成為人人推薦的教授全攻略

UrSchool系所教授評價分析

組名：我叫陳伶均請多多指教

組員：杜昕 陳伶均 廖韋凱 陳柏宇

1 專案背景與資料簡介

1.1 專案背景

還記得幾年前剛升上大學，拿到一張空白的課表，人生第一次要由自己來填上。焦急地查閱各種網路文章、評論留言、評價分數……一位位素昧平生的教授就這樣被化成一項項量化的分數，我們或多或少參考了這些分數，默默地過完了好一段大學歲月。

想想自己有沒有可能因此而與某些知識領域擦身而過，或是不小心「誤入歧途」，得要逼自己習慣不適合自己的授課方式？

從這個身為大學生揮之不去的煩惱出發，我們想知道這些對於教授、課堂的評論與分數到底是怎麼一回事。為此我們選了「Urschool大學科系教授評價網」作為研究對象，目標是找到一系列對於教授的評分與文字評價當中的端倪。

21世紀是個「線上時代」，每個人都是一個巨量數據的集合。

不論大至攸關國家社稷，小至謹謹生活瑣事，當我們面對一個陌生的問題時，第一個反應往往是上網查資料。然而透過區區幾十吋的電腦螢幕，我們能在瞬間獲得千百筆資料，新聞媒體、社群平台、電商官網……承載資料的網站五花八門，而在其間四處流竄的資料更是琳瑯滿目，在資料科學與視覺化技術的加持下，在我們眼前以各種形式呈現。

The world you see is not real, let alone what others want you to see.

S.J Watson, 《別相信任何人》

這個時代的我們就像住在一個名為「網際網路」的城市，到處閃爍著霓虹，一個不留神方向就錯了。我們需要的是一個「良好的方向感」，一個有效率獲取資訊，並正確處理和分析的素養。因此在本專案中，本組選擇了一個與身為學生的我們習習相關的主題來進行自我訓練。

1.2 專案流程

本組專案的流程包括網路爬蟲、量化分析、語意分析三項，各項流程的執行方式如下：

1. 網路爬蟲：使用 Python 爬取 UrSchool 臺大教授的個人推薦分數、同學評價等資料，並存成 Excel 檔。
2. 量化分析：使用 R 語言分析每位教授的推薦分數與所屬系所、口條、課業壓力、給分甜度等因素的關係。
3. 語意分析：使用 R 語言分析同學給教授的評價留言，分析個人推薦分數高、低的教授留言內容的差異。

1.3 資料簡介

1.3.1 讀入資料

在這次的專案裡，我們使用Python的BeautifulSoup、Selenium等套件爬取 UrSchool (<https://urschool.org/ntu/list>) 的臺大教授評價資料，包括同學對每位教授的評分與同學在教授底下的留言，並儲存成Excel檔，再使用R語言進行更進階的操作。

```
show_raw_data
```

姓名	授課系所	私心	學到東西	課業條好	給壓力	平均甜度	平均成績	最高學歷	值得一提	綽號	點名方式	
林明仁	經濟學系	教授兼系主任	5	5	5	4	4.0	A	美國芝加哥大學經濟學博士	芝加哥學派，台灣當代應用個體的代表人物	史上最帥系主任 MJ	加簽同學接受隨機抽點，其他自己選上的不點名
駱明慶	經濟學系	教授	4	4	4	3	4.0	A	美國密西根大學經濟學博士(精神不集中會很容易睡著，錯過課程精華·憤世嫉俗	NA	偶爾照點名單念一遍，原則上完全放任
孫維新	物理學系	教授	5	5	5	4	4.5	NA	美國加州大學洛杉磯分校 (UCLA) 天文學博士	2000 金鐘獎 教科文節目主持人獎	NA	NA

1.3.2 資料前處理

本組進行的前處理如下：

- 等第制資料轉換：在資料的平均成績欄位為字串形式，我們將各等地轉換為百分制。
- 授課系所統一、職稱統一：經過我們研究發現，由於UrSchool的教授名單是任何人都能添加與更新，因此資料的授課系所、職稱有各種不同的名稱，而在此我們為了方便分析而將相同意義的名稱進行整理。
- 增加學院欄位：增加各個教授的授課學院。
- 除去重複資料：有部分教授名單重複等情形，因此我們將重複的教授自資料中剔除。前處理過後的資料如下

```
show_clean_data
```

姓名	授課系所	私心	學到東西	課業條好	給壓力	平均甜度	平均成績	最高學歷	值得一提	綽號	點名方式	調整後平均成績	學院	
林明仁	經濟學系	教授	5	5	5	4	4.0	A	美國芝加哥大學經濟學博士	芝加哥學派，台灣當代應用個體的代表人物	史上最帥系主任 MJ	加簽同學接受隨機抽點，其他自己選上的不點名	87	社會科學院

駱明慶	經濟學系	4	4	4	3	4.0	A	美國密西根大學經濟學博士(精神不集中會很容易睡著，錯過課程精華. 憤世嫉俗	NA	偶爾照點名單念一遍，原則上完全放任	87	社會科學院
孫維新	物理學系	5	5	5	4	4.5	NA	美國加州大學洛杉磯分校 (UCLA) 天文學博士	2000 金鐘獎 教科文節目主持人獎	NA	NA	NA	理學院

2 量化分析

2.1 各系所各項指標分析

我們以系所為單位，比較各個學院私心推薦、學到東西、口條好、課業壓力、給分甜度五項指標的平均，結果如下

depart_score_analysis

學院	教授資料數	平均私心推薦	平均學到東西	平均口條好	平均課業壓力	平均給分甜度
文學院	271	4.47	4.32	4.21	3.21	3.96
理學院	169	4.10	4.03	3.90	3.41	3.78
社會科學院	130	4.24	4.12	3.94	3.05	3.72
醫學院	112	4.00	4.21	3.96	3.41	3.67
工學院	154	4.19	4.14	3.94	3.29	3.96
生農學院	177	3.80	3.75	3.68	3.11	3.72
管理學院	118	4.19	4.15	3.99	3.12	3.94
公衛學院	16	4.25	3.91	3.78	3.25	3.69
電資學院	111	4.29	4.25	4.06	3.59	3.93
法律學院	48	4.53	4.56	4.55	3.47	4.23
生命科學院	53	3.79	3.95	3.73	3.11	3.58
共同教育中心	1	4.00	4.00	4.00	4.00	3.00
其他	3	4.50	4.67	4.67	4.50	3.67

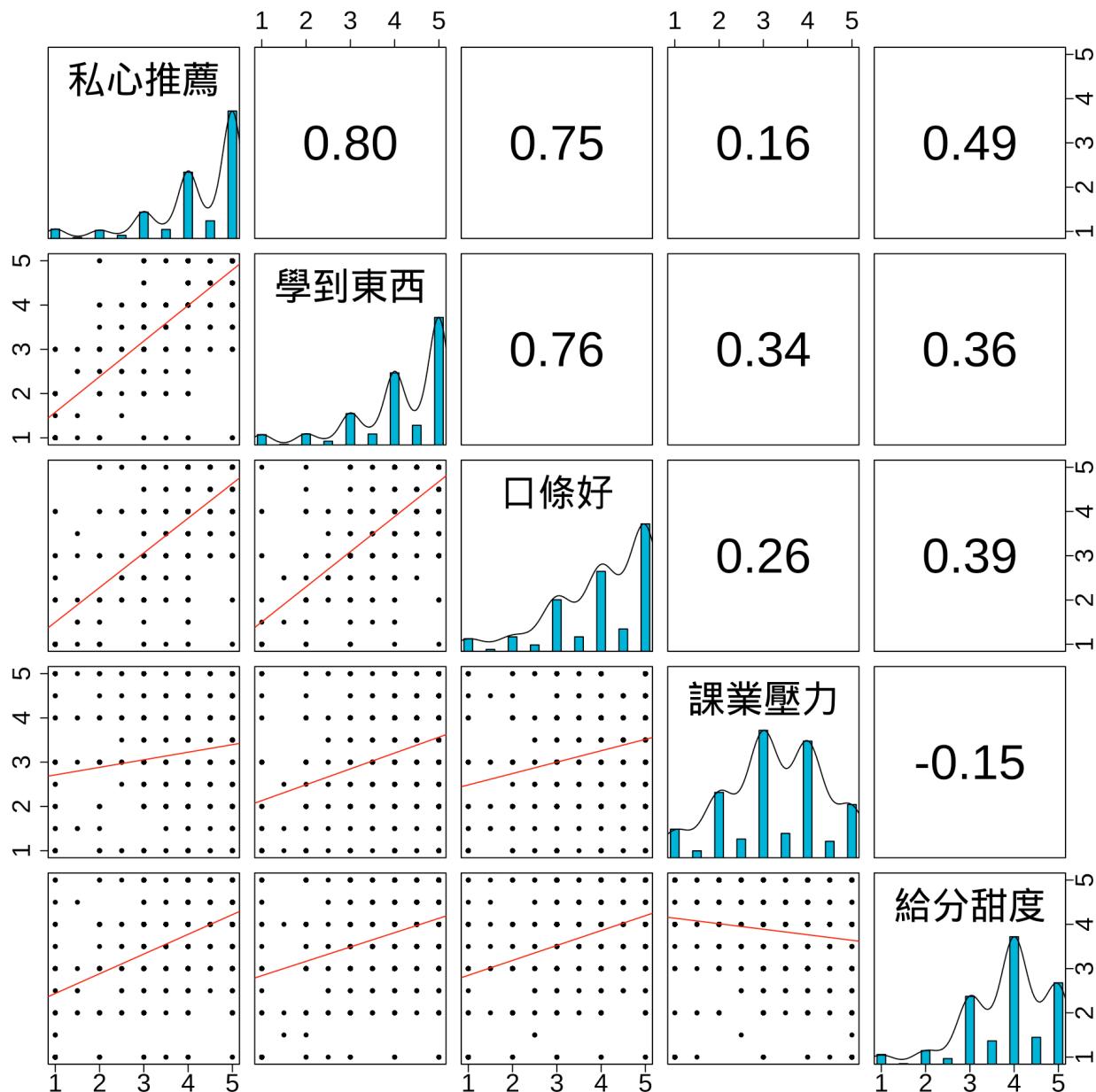
由上表我們可以發現以下幾項資訊：

1. 文學院的資料筆數最多，可能是因為資料中有許多大學國文的老師及教授外語的老師。
2. 法律學院在平均私心推薦、學到東西、口條、給分甜度分數都是所有學院最高的。可能代表法律學院教授在各項指標較獲得學生喜愛，或除了教授給分比較甜外，評分的學生給分也比較甜。
3. 生農學院、生命科學院在平均私心推薦、學到東西、口條、給分甜度分數都是偏低的，皆未超過4分。
4. 平均課業壓力以電資學院、法學院最高，而社會科學院最低。

2.2 各項指標與私心推薦分數的相關性

分析完各個系所各項指標的差異之後，接下來分析各項指標之間的相關性。如下圖，可發現同學私人推薦分數與是否學習到東西及教授的口條如何呈現高度正相關，其次為給分甜度，而相較之下課業壓力與私心推薦分數的相關性最低。我們認為是因為每個人偏好的課業壓力不同，且課業壓力高、低都有可能會受到學生的推薦，但應該不會有同學會推薦學不到東西或口條不佳的教授。而我們原本預想給分甜度的相關性可能與學到東西、口條好等指標相近，結果實際上差了一段距離，代表各位臺大的學生比起課程甜不甜，更重視是否能學習到知識，及上課的體驗，實在非常欣慰。

```
scores %>%
  pairs.panels(lm = T, smooth = F, ellipses = F, rug = F, hist.col = "#00b4d8", cex.labels=3, cex.axis=2)
```



2.3 使用迴歸模型預測推薦分數

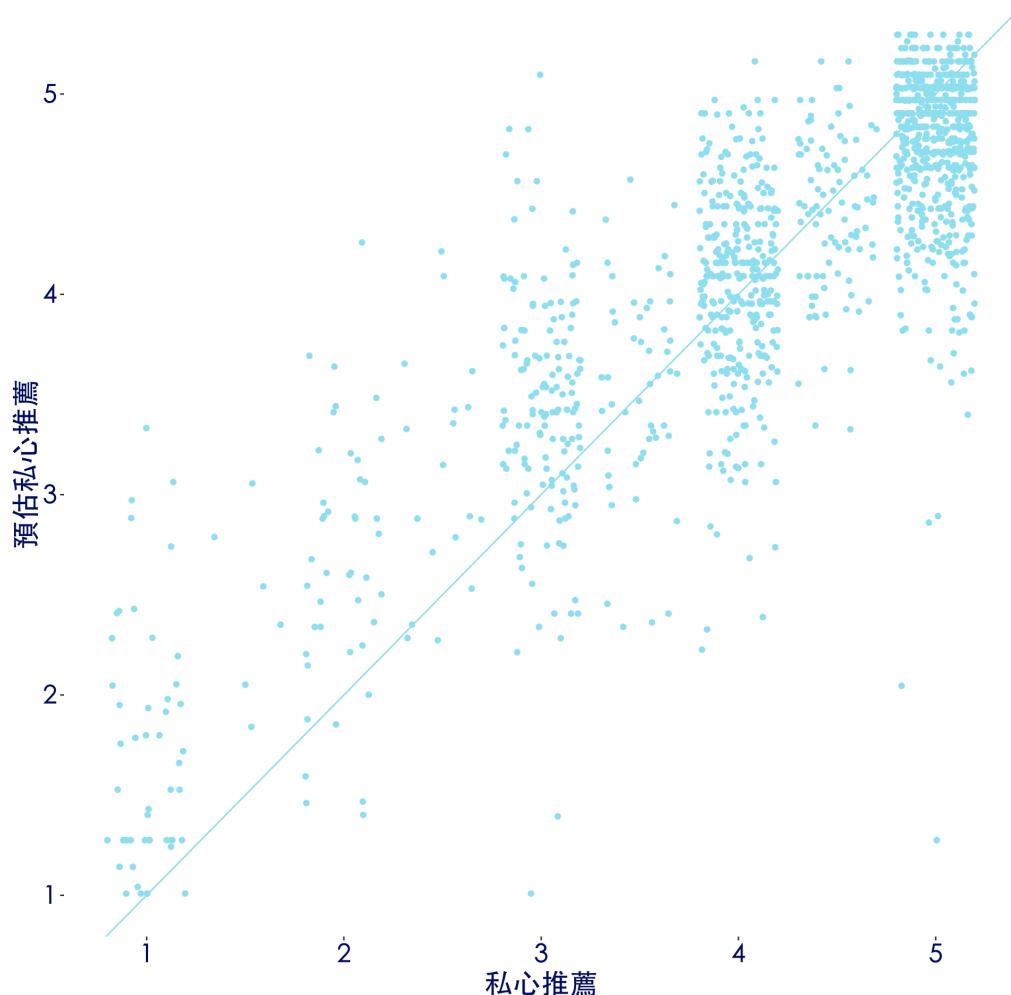
除了分析各項指標與私心推薦分數的相關性外，我們還使用傳統複迴歸模型求出私心推薦分數與各項指標的線性關係，如下

regression_result

```
##  
## Call:  
## lm(formula = 私心推薦 ~ 學到東西 + 口條好 + 課業壓力 +  
##      級分甜度, data = data)  
##  
## Coefficients:  
## (Intercept)    學到東西    口條好    課業壓力    級分甜度  
##       0.33648     0.54096     0.27173    -0.06662     0.19257
```

得到迴歸模型的各項係數後，我們使用此係數估計每位教授的私心推薦分數，並比較其與實際私心推薦分數的關係，繪製成散布圖，如下

predict_result



2.4 其他可能影響私心推薦分數的指標

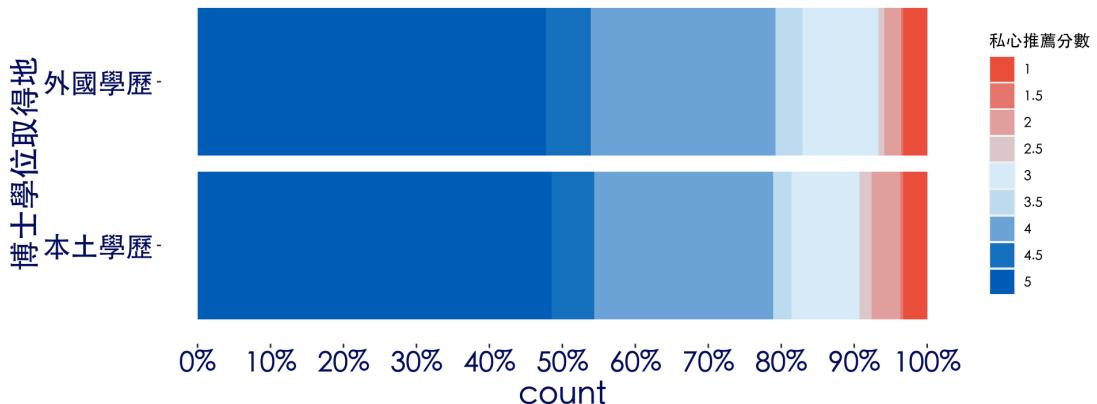
除了原始資料已經有的數量資訊外，我們針對基本資料中的其他文字欄位進行分析，判斷是否與私心推薦分數有關聯。

2.4.1 土博士 v.s. 洋墨水

是否留學海外的教授可以在私心推薦分數上擁有較好的表現？以下我們透過檢查資料的「最高學歷」項目，將最高學歷於國內學校獲得者歸類為「本土學歷」，反之則歸類為「外國學歷」。完成歸類後，檢視私心推薦分數表現與之是否有關聯。然而由於網路資料在「最高學歷」項目的名稱上相當多元，透過以下篩選將有部分誤差。本部分結論有二：

1. 依照此歸類資料，外國學歷資料數多於本土學歷（可能包含未被篩選出的本土學歷），因此即使高分群以擁有外國學歷的教授佔多數，也不一定代表兩個項目有因果關係或相關性。
2. 不論在本土學歷或是外國學歷項目中，私心推薦分數獲得5分的教授皆為最多，或許意味著兩者相關性並不強。

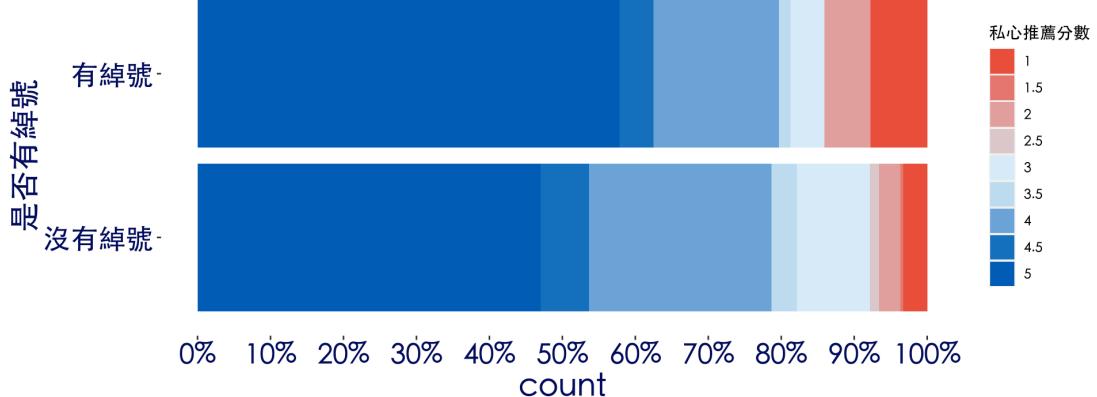
phd_place



2.4.2 教授有綽號是否代表與學生比較喜歡？

以下驗證有無綽號與私心推薦分數表現的關聯性，發現雖然有綽號的教授僅佔約2%，但觀察有綽號的教授在各分數階層的分布時可以發現，有綽號的教授私心推薦分數為5分的比例超過50%，高於沒有綽號的教授的比例，但相對的有綽號的教授私心推薦分數為1分、2分的比例也較沒有綽號的教授高，可推測教授可能在學生推薦或不推薦一名教授的時候，較可能為該位教授取綽號，而綽號也可能因此為褒義或貶義。

nickname



3.2 學生最重視什麼？（各個學院）

對各學院的教授評論斷詞，觀察各學院學生最常評論教授的什麼特質。由下表可見，各個學院的學生之間所重視的特質幾乎相同，幾乎都是有趣、認真、清楚，值得注意的是，公衛學院的最高出現形容詞為無聊，出現次數僅有兩次，推斷是因為公衛學院的教授幾乎都無評論。

important_department

學院	最常形容	出現次數
文學院	有趣	120
理學院	認真	75
社會科學院	清楚	51
醫學院	清楚	18
工學院	認真	43
生農學院	認真	42
管理學院	認真	33
公衛學院	無聊	2
電資學院	清楚	30
法律學院	清楚	13
生命科學院	有趣	8
共同教育中心	NA	NA
其他	有趣	2

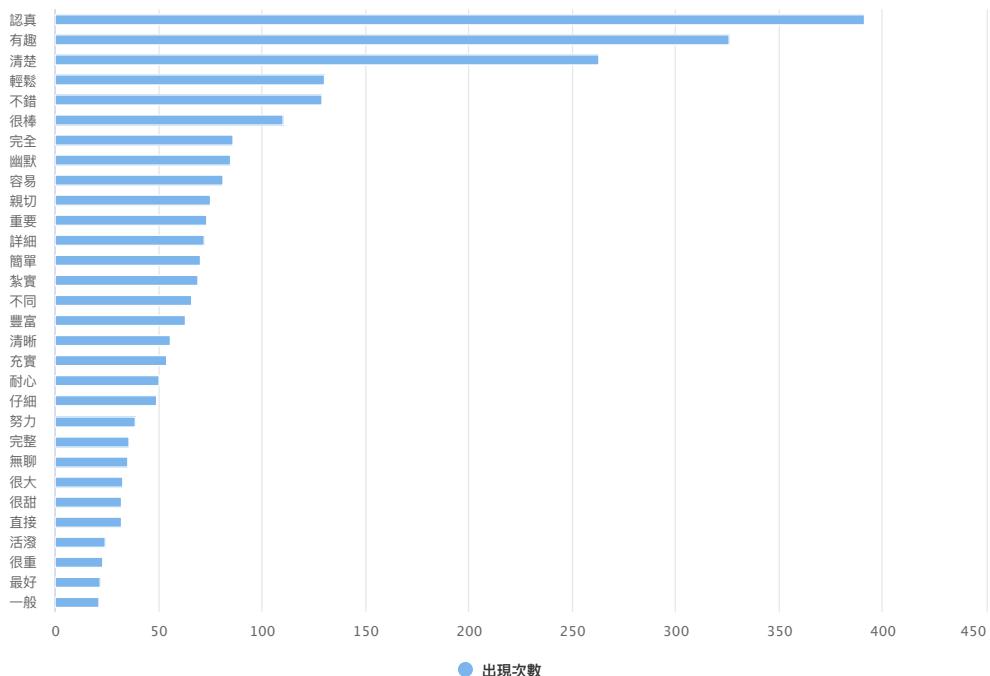
3 語意分析

3.1 學生最重視什麼？

從評論中利用jieba斷詞，並取出形容詞，觀察評論中出現最多的形容詞為何，就可推導出學生最重視什麼特質，分別做成圖表與文字雲如下，可見學生最重視的三大教授特質為認真、有趣、清楚。

important_personality

評論中最常出現的詞彙 Top 30



wordcloud

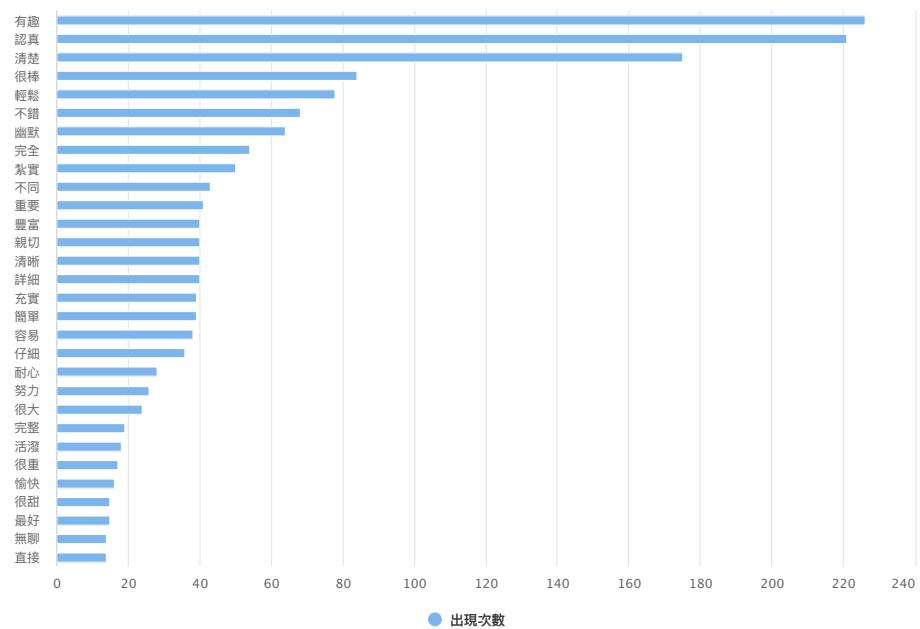


3.3 高評價教授 v.s. 低評價教授

對高評價教授和低評價教授的評論斷詞，觀察兩者的評論內容最常出現的形容詞有何不同。由下表可見，在此我們將所有私心推薦分數為5分的教授列為高評價教授，共433位。而低評價教授則為私心推薦分數小於3分的教授，共130位。高評價的教授的評論幾乎都是正面評價，而且正面字眼的出現頻率很高；相反的，低評價的教授也有正面評價，但出現頻率低很多，反而出現了很多負面字眼。

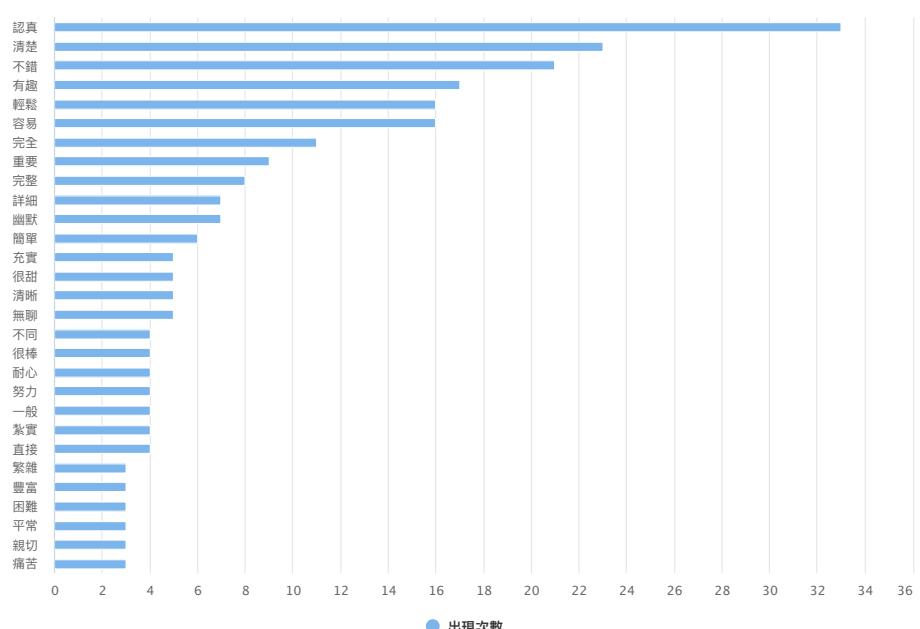
`important_personality_high_score`

高評價教授評價最常出現的詞彙 Top 30



`important_personality_low_score`

低評價教授評價最常出現的詞彙 Top 30



4 總結

綜合上述分析，我們判斷學生選課時可使用以下兩個項目來進行決策：

1. 各項指標分數
2. 評論內容

首先，各項指標分數的部分，可以先從私心推薦分數對於教授和課堂獲得最粗略的認識，接著再以學到東西、口條好、給分甜度來判斷較細節的部分。至於課業壓力分數則是見仁見智，對推薦分數沒有什麼影響。

其次，評論的部分可以參考留言者的用詞。我們發現私心推薦分數高的教授，相關的評論幾乎都是正面的評價，沒有負面的字眼；但私心推薦分數低的教授，相關的評論雖然還是會出現正面評價，但負面字眼出現的頻率也不低。

最終結論：選課時，可先從各項分數了解大致的情況，再參考同學的評論，來判斷該教授的教學情況。

附錄 分工表

杜昕：資料前處理、量化分析、網頁架設、簡報製作與表達

陳怜均：網路爬蟲、語意分析

廖韋凱：網路爬蟲、語意分析

陳柏宇：量化分析、網頁文字撰寫