

2021 電子商務技術 作業二

第一部分：匯入 pandas 及所需的繪圖套件(matplotlib、seaborn 等)。請依題號將程式碼與輸出結果截圖到答案卷中。

1. 載入資料集 Advertising.csv 並將 Date 指定為 DataFrame 的 index，請列出前五筆資料，如圖 1 所示。(2%)

	TV	newspaper	radio	sales
Date				
2018-01-01	230100.0	69200.0	37800.0	22100.0
2018-01-02	44500.0	45100.0	39300.0	10400.0
2018-01-03	17200.0	69300.0	45900.0	9300.0
2018-01-04	151500.0	58500.0	41300.0	18500.0
2018-01-05	180800.0	58400.0	10800.0	12900.0

圖 1

2. 請使用 pandas 協助了解此資料集，並根據輸出結果回答下列問題。
 - a. 各欄位的資料型態為何 (2%)
 - b. 是否有空值，如果有的話請列出有空值的欄位 (2%)
3. 請使用 pandas 協助了解資料的分布狀況 (pandas 提供許多方便的統計函數)
 - a. 請計算各欄位的資料筆數(count)、平均值、標準差、最小值、最大值、第 1/2/3 四分位數。(2%)
 - b. 請計算各欄位的第 10、20、30、40、50、60、70、80、90、100 百分位數。如圖 2 所示。(2%)

	TV	newspaper	radio	sales
0.1	24880.0	5990.0	3400.0	7960.0
0.2	59180.0	9380.0	7680.0	9700.0
0.3	86750.0	15970.0	11940.0	10870.0
0.4	119000.0	21320.0	17120.0	11800.0
0.5	149750.0	25750.0	22900.0	12900.0
0.6	186060.0	32700.0	27860.0	14800.0
0.7	210730.0	39110.0	33500.0	16600.0
0.8	228540.0	49640.0	38920.0	18920.0
0.9	261440.0	59070.0	43520.0	21710.0
1.0	296400.0	114000.0	49600.0	27000.0

圖 2

4. 請繪製各欄位的直方圖(一個欄位一張，不限格式)。(5%)
5. Scatter Matrix 可以展現出兩兩變數之間的關係，請繪製 Scatter Matrix (10%)，並依據此圖回答 TV、newspaper、radio 這三種媒體的觸及數和 sales 之間的關係為何。(5%)
6. 請依序完成下列題目，利用 pandas 將每日資料統整成每月資料並將此數據視覺化。
- a. 請將 index 的資料類型轉成日期或時間(提示：需查看在 pandas 支援何種日期、時間格式) (5%)
- b. 請將每日資料統整成每月資料且 index 格式為「年/月」，如下圖所示。(5%)

	TV	newspaper	radio	sales
2018/1	4404000.0	1173100.0	753900.0	424900.0
2018/2	4693800.0	824900.0	696600.0	434200.0
2018/3	3909100.0	923200.0	799900.0	428900.0
2018/4	4823300.0	1022600.0	649700.0	440700.0
2018/5	3973700.0	926200.0	800300.0	410000.0
2018/6	4674400.0	834200.0	573000.0	395800.0
2018/7	2930200.0	406600.0	379400.0	270000.0

- c. 為了解各媒體的觸及數占比，請繪製每月各媒體的圓餅圖，需有 title、legend、百分比資訊。(5%)
- d. 為方便比較每月各媒體觸及數變化情況，請繪製 100% stacked bar chart。(10%)
- e. 為了解 sales 每月變化，請繪製 sales 的折線圖。(5%)
- f. 為看出各媒體觸及數實際數值，請在同一張圖上繪製 3 種媒體的折線圖，需有 title、legend、x 軸/y 軸標籤。(5%)

第二部分：請載入 sales.csv 和 sales2007.csv，並利用 pandas 完成以下題目。請依題號將程式碼與輸出結果截圖到答案卷中。

1. 請將 sales.csv 與 sales2007.csv 合併成一個 DataFrame。(5%)

(以下題目皆使用合併後的 dataframe)

2. 請列出此資料集中每個 Product line 的出現次數。(5%)
3. 請加總每個國家的 Revenue (2%)，請問哪些國家的 revenue 大於 30,000,000。(3%)
4. 請加總每年的每個 Product 的 Revenue、Planned revenue 等資料類型為數值的欄位，產出一個 Multiindex dataframe。(第一層 index 為 Year，第二層 index 為 Product) (5%)
5. 請從上一題產出的 DataFrame 中取出每一年銷售量(Quantity)最多的產品。(10%)
6. 請新增一個欄位「Gross margin」，並填入銷售毛利率值。(5%)

-
- 繳交期限：3/31(三) 中午 12:00
 - 請繳交答案卷(.pdf)和程式檔(.ipynb)，檔名為 ECT_HW2_學號
 - 上傳至 ee-class 作業區，遲交一天扣該次作業 5%