

DS 108 Homework2

Notes for Grading

datascience2019fall@gmail.com

Rules

1. 這次作業請大家以 Report (PDF檔)的方式繳交 (格式不對扣10%)

學號(小寫).zip

----學號(小寫)_hw2.pdf

Ex: r12345678.zip

----r12345678_hw2.pdf

2. No Plagiarism

Problem 1: (Extract data and handling of CSV files and RDB) 20%

1. 請同學在報告中讀完 2019年6月到 2019年10月的csv檔 (2019.06.01 到 2019.10.09 不包含 10月14,15日), 用 pandas 將讀到的資料 (dataframe的 head()) 和統計數據 (dataframe的describe()) 印出來, 僅需處理在證交所上的資料, 不須額外補數據。
2. 請同學在報告中寫出是用什麼指令將 dataframe 寫進 MySQL Database

Problem 2: (Basic calculation) 20%

1. A小題請同學寫下(或是截圖) 你們的Python Program, 是用什麼SQL指令算出總數以及平均數, 成交總金額和交易總股數都要, 共四組答案(指令,答案)
2. B小題請同學寫下你們 Pandas 指令以及印出你們兩組的值

A: From your python program issue SQL statement to find out:

- Total and average trade amount in this time period(成交總金額)
- Total and average trade volume in this period (交易總股數)

B: Use pandas to calculate

- 5-day moving average of daily closing price.
- 5-day moving average of daily trading volume.

Problem 3: (Simple vector calculation) 20% Find the biggest daily change.

1. 請同學在報告中寫出題目要求的日期和數值 (四組答案), 另外請同學在報告中說明是下什麼指令做計算的, by percentage 是以前一天為基準, 成長了多少%

By absolute amount:

- a. Highest daily price
- b. Daily trade volume

By percentage:

- c. Highest daily price
- d. Daily trade volume

Problem 4: (Outlier) 20% Special days

1. 印出五個 outlier 的 日期, 三個維度一起算, 算距離請使用 L2

Use [daily volume], [closing price - opening price], [highest price - lowest price] as your three axis, normalized by their average.

Find the top five days that are most far away from the average.

Problem 5: (Most profit) 20%

1. 請使用收盤價
2. 答案不應該有交易停止日(包含假日)
3. 請在報告中寫出 TOP 5 獲益最高的 Triples (買入日期, 賣出日期, 獲益)
4. 請同學在報告中說明是用何種演算法做計算的

Deadline

截止日期: 2019/10/23 09:00 a.m.

遲交期限: 2019/10/25 09:00 a.m. (遲交一天扣20%)