Data Structures
Final Project
Due: June/21/2024

In this final project, you have to solve two major problems.
分數配分比例如下:

Problem 1: 40%
Problem 2: 32%
Coding style: 18%
Report: 10%

Each group has to compress the relevant files for this project and submit the compressed file onto Moodle. Each group's compressed file has to contain

(1) All the source codes
(2) The final report (no more than 20 pages in word file). In this report, you have to describe how you solve the problem and what results (可用截圖呈現) you obtain for each question. In the report, also make sure to describe each student's role in the final project.

In this project, when you have to find the maximum, minimum and median values, the requirement is that you have to implement C++ classes for Heapsort and use these classes for sorting (you have to write C++ classes, rather than just using C language for this project) (實做出 heapsort 的 C++ classes 並用其做 sorting 為此專題之基本要求). Additionally, you may use other libraries to solve the problem for comparison (可額外加分).

Problem 1: Single financial product problem -- daily prices of TWII (TSEC weighted index, 臺灣加權股價指數)

Dataset: "TWII_withRepeatedData.xlsx"

The dataset contains data from 5 columns as follows:

Column 1	Column 2	Column 3	Column 4	Column 5
Date	Open_price	High_price	Low_price	Close_price

Please write C++ codes for the following tasks:

Task (A): (共 20%, 底下(1)至(10)每小題各 2 分)

(1) Determine how many unique dates are in the dataset.

After removing the data from repeated dates, please use the closing price (i.e., the column for "Close_price") to solve (2) – (9):

(2) Find the 10 smallest prices and which dates contain these smallest prices.

(3) Find the 10 largest prices and which dates contain these largest prices.

(4) Find the median price and its occurring date

(5) Compute the daily return for every day (except the first day). Then determine what the maximum and minimum returns (return could be a negative value) are and on which day(s) they occur.

Daily returns is defined as:

[(P(t+1)-P(t))/P(t)]*100%, where t is date.

(6) Compute the intraday return for every day. Then determine what the maximum and minimum returns (return could be a negative value) are and on which day(s) they occur.

Intraday returns is defined as:

[(Close_price(t)-Open_price(t))/Open_price(t)]*100%, where t is date.

(7) Make a plot of the closing price over time, in which the x-axis is the day index and y-axis is the price.

(8) Make a plot of the daily return over time, in which the x-axis is the day index and y-axis is the daily return.

(9) Make a plot of the intraday return over time, in which the x-axis is the day index and y-axis is the intraday return.

(10) Find the maximum, minimum and median prices using all the 4 columns of prices (i.e., Open_price, High_price, Low_price and Close_price) and determine on which date they occur.

Task (B): (共 20%, (1)至(10)每小題各 2 分)

For the same dataset used in Task (A), after removing the data from repeated dates, please generate a set of sampled data at the interval of every five days, then do the same tasks for (1)-(10) of Task (A).

Problem 2: Multiple financial products problem -- tick-based price data of options (選擇權)

Datasets:

"OptionsDaily_2017_05_15.csv"
"OptionsDaily_2017_05_16.csv"
"OptionsDaily_2017_05_17.csv"
"OptionsDaily_2017_05_18.csv"
"OptionsDaily_2017_05_19.csv"

Each dataset contains data from 9 columns as follows:

Column 1: 成交日期
Column 2: 商品代號
Column 3: 履約價格
Column 4: 到期月份(週別)
Column 5: 買賣權別
Column 6: 成交時間
Column 7: 成交價格
Column 8: 成交數量(B or S)
Column 9: 開盤集合競價

A financial product is defined by a unique combination of columns 2 to column 5. I.e., if a value generated by combining the values of column 2, 3, 4, and 5 is unique, then it defines a unique product. For example, in "**OptionsDaily_2017_05_17.csv", the 3rd row contains the following data:**

Column 2 = CAO
Column 3 = 70

Column 4 = 201706
Column 5 = P

So the product “CAO_70_201706_P” is a unique options product.

Please write C++ codes by OOP coding style for the following tasks:

(1) Determine, totally, how many unique products exist in all these five datasets. (3%)
(2) Determine if the product “TXO_1000_201706_P“ exists in these datasets. (3%)
(3) Determine if the product “TXO_9500_201706_C“ exists in these datasets. (3%)
(4) Determine if the product “GIO_5500_201706_C“ exists in these datasets. (3%)
(5) For “TXO_9900_201705_C“,
  a. Find the 10 smallest prices and what time (determined by the unique combination of Column 1 (成交日期) and Column 6 (成交時間)) these smallest prices show up. (5%)
  b. Find the 10 largest prices and what time these largest prices show up. (5%)
  c. Find the median price of this product. (5%)
  d. Compute the ticked-based return (except the first tick) using the values in Column 7 (成交價格). Then determine what the maximum and minimum returns are and when they occur. (5%)