

Philippine Stock Market Data: (some simple) Analysis



Note: logo images obtained from PSE website and some of the companies

Figure 1. Logos of the 30 companies in the Philippine Stock Exchange Index (PSEI)

I. INTRODUCTION

Big Data and *Data Science* are recent buzz words in multiple disciplines including the sciences and engineering. For this machine problem, we will use actual data generated from the buying/selling of publicly listed stocks in the Philippine Stock Market. More specifically, we will use data for the 30 companies in the Philippine Stock Exchange Index (PSEI). The logos of these companies are shown in Figure 1. How many logos look familiar to you?

Concepts covered in CCPROG2, i.e., arrays, string, structures, text and binary file processing will be applied.

Through this MP, demonstrate that you can:

- perform data gathering
- design and implement your own data structure for representing, storing, accessing and manipulating stock historical data
- design and implement your own algorithms for the challenge problems described in the next section
- specify test cases
- test and debug programs
- properly document and articulate your solution to the MP

II. CHALLENGES [Note: Each Challenge is worth 10 points unless explicitly stated otherwise]

Challenge #1: Gather Stock Historical Data (SHD) [5 points only].

As a preliminary task, we need to gather **historical data** for stocks in the Philippine Stock Exchange Index (PSEI)¹. This is an easy task that can be done in less than 30 minutes since the data that we need are downloadable (for free) from the Wall Street Journal website. For example, try downloading data for Ayala Corporation from <http://quotes.wsj.com/PH/AC/historical-prices>.

Do this by setting the start date to **1/1/2014** and end date to **12/30/2018**, and then press the “GO” button. Thereafter, press the “DOWNLOAD A SPREADSHEET” button to download the data into an Excel file. A sample screen capture is shown in Figure 2 for your reference.

Open the Excel file. As shown in Figures 3 and 4, it contains data in six columns, namely: Date, Open, High, Low, Close and Volume². The Open, High, Low, Close are abbreviated as OHLC.

Take note of the last row number that contains data. In the AC example, that row number is 1216.

Select all cells (press Ctrl + A), copy (press Ctrl + C) and paste (press Ctrl + V) **to a new text file**. Close the Excel file, you will no longer need it.

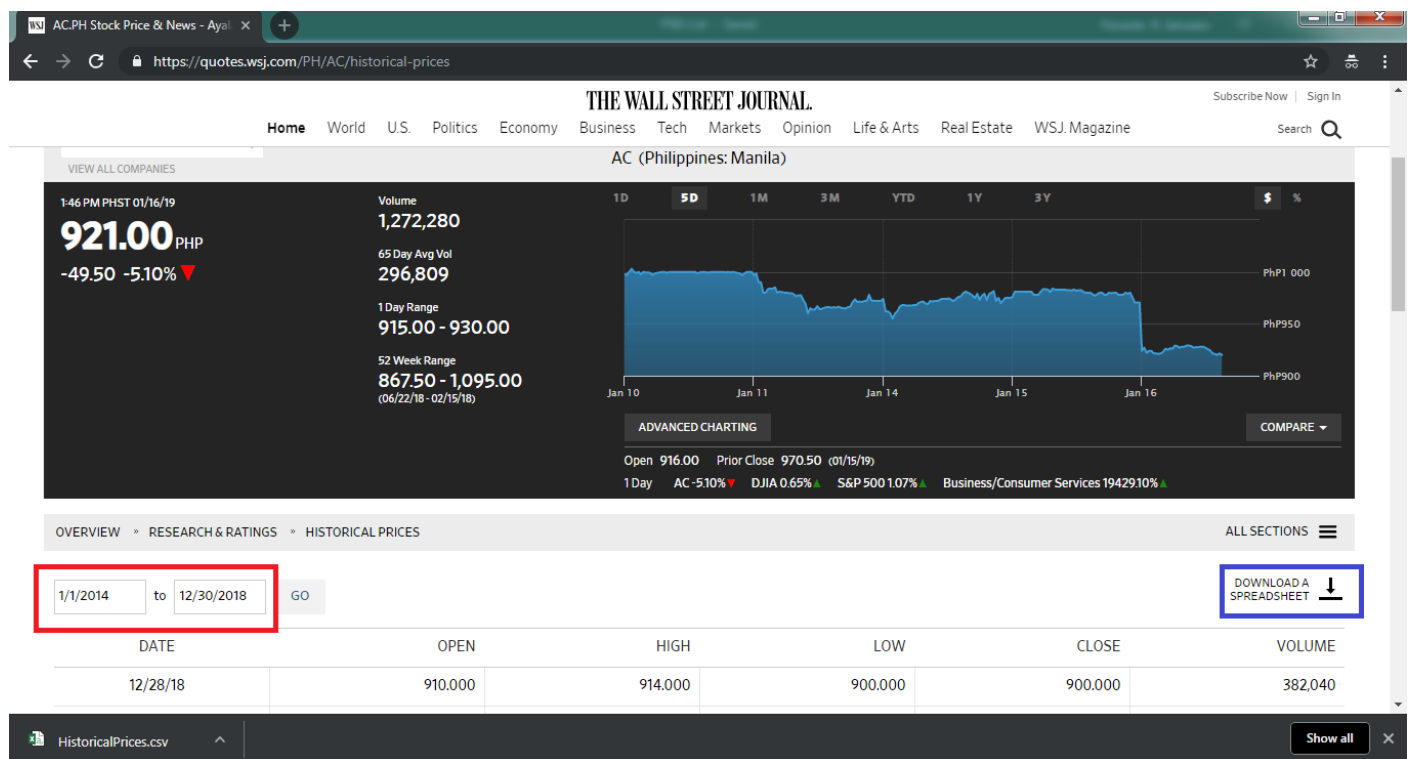


Figure 2. Screenshot for downloading AC stock historical data from 1/1/2014 to 12/30/2018.

¹ <http://www.pse.com.ph/stockMarket/marketInfo-marketActivity.html?tab=1#>

² “Open” is the price on the very 1st trade of the day, while “Close” is the price at the end of the day. “High” and “Low” are the highest and lowest prices of the stock respectively for the day. “Volume” is the number of trades.

	A	B	C	D	E	F	G	H	I	J	K
1	Date	Open	High	Low	Close	Volume					
2	12/28/2018	910	914	900	900	382040					
3	12/27/2018	916	924	905	905	274520					
4	12/26/2018	918	918	909	911	133400					
5	12/21/2018	918	919.5	910	914	404820					
6	12/20/2018	914	925	906	918	414360					
7	12/19/2018	912	915	907	914	462640					
8	12/18/2018	920	925	916	918.5	257640					
9	12/17/2018	915	930	915	927	367430					
10	12/14/2018	920.5	940	905.5	914	224510					

Figure 3. Screenshot for AC stock historical data.

	A	B	C	D	E	F
1213	1/7/2014	529.5	529.5	523	524	197230
1214	1/6/2014	525	530	519	525.5	399380
1215	1/3/2014	521	523.5	514	520.5	552230
1216	1/2/2014	520	528.5	519	525.5	367200
1217						

Figure 4. Row 1216 contains the last row of daily transaction data.

Make sure you're now in the new text file with the stock data; do the following. Delete the first row (i.e., the line containing the words Date Open High Low Close and Volume). Replace it with a line that contains the stock symbol, followed by at least one space, followed by a number indicating the number of daily transactions. The number of daily transactions is the last row number in the Excel file minus 1. In the AC example, it corresponds to 1215. Thereafter, insert one empty line before the first daily transaction data. Save the text file using the concatenation of the stock symbol and "orig" as filename. For example, in the case of AC data, the text file should be stored as **ACorig.txt** (Note: you will receive a copy of ACorig.txt file together with this MP specification document; open and examine its contents).

Each student will be assigned to gather historical data for some stocks. Refer to the tables below for your individual stock assignment. Look for your ID number. The stock assigned to you is indicated in the cell right next to your ID number.

Just copy and paste <http://quotes.wsj.com/PH/AC/historical-prices> on your browser, then replace AC with the stock code assigned to you, and proceed as described in the previous paragraphs.

The number of daily transactions is not the same for all stocks! You will need to check the last row with data in the Excel file.

DELIVERABLE: submit the text file for the stock assigned to you via Canvas. Take note of the Canvas submission deadline.

Note: the text files will be collated and provided to you for the remaining challenges. Since there are 30 companies in the PSEI, there will be 30 text files.

Stock Assignment Table

11827467	AEV
11825391	AGI
11710454	ALI
11828897	AP
11712074	BDO
11829109	BPI
11827696	DMC
11827335	FGEN
11827734	GLO
11845708	GTCAP
11843470	ICT
11811838	JFC
11800267	JGS
11800348	LTG
11828714	MBT
11831715	MEG
11833076	MER
11715421	MPI
11730064	PCOR
11830077	PGOLD

11827467	RLC
11825391	RRHI
11710454	SCC
11828897	SECB
11712074	SM
11829109	SMC
11827696	SMPH
11827734	TEL
11845708	URC
11843470	AEV
11811838	AGI
11800267	ALI
11800348	AP
11828714	BDO
11831715	BPI
11833076	DMC
11715421	FGEN
11730064	GLO
11830077	GTCAP

11827041	ICT
11827025	JFC
11826568	JGS
11838698	LTG
11827823	MBT
11827750	MEG
11829087	MER
11837012	MPI
11829001	PCOR
11828528	PGOLD
11824875	RLC
11829303	RRHI
11827289	SCC
11828757	SECB
11617365	SM
11849088	SMC
11829591	SMPH
11829494	TEL
11829923	URC

Challenge #2: Make the Format Uniform.

The downloaded data stored in the text files do NOT actually follow a uniform format. For example, data for MER has different formats for date – in particular, year is coded as either 18 or 2018 (see Figure 5).

MER 1215

12/28/18	394.4	394.4	380	380	280020
12/27/18	389.4	394.8	386	394.8	254420
12/26/18	385.4	389.6	380	389.6	97320
12/21/18	390	391	385.4	390	313870
12/20/18	389	390	386	390	186580
12/19/18	389	389.8	387.2	389.4	221670
12/18/18	389	389	386.2	389	222960
12/17/18	387.2	390	380	386	135170
12/14/18	375.2	389.8	373.2	389.8	280430
12/13/18	386.2	387.4	375	375	414550
12/12/2018	377	382.4	373	382.4	192640
12/11/2018	373	378.4	369	377	346680
12/10/2018	375.4	375.4	366.2	366.2	160030
12/07/2018	380	381	375.4	376	154800
12/06/2018	385.2	386	380	380	239940

Figure 5. Dates for MER do not follow a consistent format (see year)!

Your task is to write a program that will copy the contents of each text file into another text file. There are two differences though:

1. The destination files should contain dates in uniform format following MM/DD/YYYY. For example, if the original data is "1/1/2014", then the new date in the destination file should be "01/01/2014". Another example, if the original date is "1/3/18", then the new date should be "01/03/2018".
2. The values for OHLC and Volume should be encoded as floating point values with two digits after the decimal point.

It is up to you to decide how many white spaces you want to put to separate two data values within the same line of text.

Your program should not ask for any input, or display anything on the screen. Your program should produce 30 new text files. Each destination file should have a filename corresponding to the stock code. For example, if the original source file is "ACorig.txt", then the destination file should be named as "AC.txt".

DELIVERABLE: submit your C source code via Canvas. Take note of the Canvas submission deadline.

Challenge #3: Design and Code Your Own SHD Data Structure.

Design and code the data structure for representing and storing the stock names and historical data in the primary memory. Take note again that there are 30 companies comprising the Philippine Stock Exchange Index. Your data structure design should be able to represent and store the 5 year stock historical data for all the 30 companies! The following are at your disposal: arrays, strings, and structures.

You are required to use the following pre-declared data types which are encoded in "CCPROG2_MP.h" header file (*note: you will receive a copy of this header file*). The **struct DailyData** is the data structure for representing and storing a single day data.

```
/* CCPROG2_MP.h */

/* DO NOT CHANGE the contents of this file */

#ifndef CCPROG2_MP_DEF
#define CCPROG2_MP_DEF 1

typedef char StrDate[11];    // for dates in string format "MM/DD/YYYY", ex. "10/24/2016"
typedef char Symbol[9];     // for stock names, ex. "EEI", "SMPH", "URC"

struct DailyData {
    StrDate date;
    float open, high, low, close; // OHLC
    float volume;
};

typedef struct DailyData DDType;

#endif
```

Your output for this challenge is very important because your data structure will affect the algorithms that you will implement in the succeeding challenges.

DELIVERABLES: submit the two files described below via Canvas. Take note of the Canvas submission deadline.

1. You need to submit a softcopy of your header file containing the declaration/s for your data structure. Encode your submission using the template **LASTNAME_DS.h** file. Use your lastname as filename. For example, if your last name is SANTOS, then your file submission should be **SANTOS_DS.h**. By the way, DS means Data Structure.
2. You also need to submit a PDF file (corresponding to at most 4 pages of short bond paper, stapled) containing the header file codes, and an explanation/discussion of your data structure. Make sure to include a figure/diagram to help me understand your design. Use your lastname as filename; example: **SANTOS.PDF**.

To be continued... (Part 2 will be sent in another email).