CS 111: Introduction to Computer Science **Winter 2022**
Problem Set #7: Minnesota Mining and Manufacturing

How low can you go?
Can you go down low?
All the way to the floor?
How low can you go?
Can you bring it to the top?

Willie Perry, Jr. (D.J. Casper) (b. 1971)
"Casper Slide Pt. 1" (2000)

This is a pair assignment, to be done in your newly assigned pairs. *Remember how pair programming works:* you and your partner are to do all of the work together, at the same time, at the same computer. Two brains, one keyboard. Problems #1, #2, and #3 are required; problems #4 and #5 are totally optional. As always, when you have questions, feel free to consult David, Chloe or Aishwarya, the book, the lab assistants, or to seek consulting help from a classmate.

This assignment is intended to give you a chance to do some data processing, and, along the way, get a little practice with loops and string processing. You'll be given a file of data from 3M (purveyors of Post-It Notes™, masks, a thin-film material that covers the back-up camera on your car to help keep the lens clean throughout an entire Minnesota winter, etc.), and you'll be asked to answer a few basic questions about it.

1. Download the data from Moodle. The lines of the file are of the following form:

   ```
   MM/DD/YYYY,Close,Volume,Open,High,Low
   ```

   `Open`, `High`, `Low`, and `Close` are all stock prices (for what price could you buy a single share of 3M stock? including the first, highest, lowest, and last prices of the day). The `Volume` is the number of shares that were bought and sold that day.

   Write a program to load this dataset (take a look at previous problem sets for some help; this will look very similar to the Capital Bikeshare assignment). You'll need to open the file, then read it line by line (splitting on commas to divide it into separate fields), then close the file at the end. You'll need to figure out a way to remove the `$` from the various stock prices. (Maybe `replace` would be helpful to look up in Zelle? Or you can do it with string indexing.)

2. Compute the number of days in 2021 in the dataset in which the volume was more than 5,000,000. *(Hint: use* `split` *twice.)* There are many lines from the file that contain data from other years, so you will need to figure out which lines should count. Print out the result in a reasonable sentence.

3. Compute the average high, average low, average opening, and average closing price in the year 2021. Print out the result in a reasonable sentence.

4. Compute the number of days in 2021 in which the opening price was at least $1 higher than the previous day's opening price. You may assume that the datafile is sorted from the beginning of time to the end of time. (One way to do this: create a variable that stores the previous line's opening price, and compares the two values. You'll need to do something special to make sure that your program does the right thing on the first day you encounter, when there *isn't* a previous opening price.) Print out the result in a reasonable sentence.

5. I like my investments to be boring. I hope you do too; there's really good evidence that investing for the long haul (with a broadly diversified portfolio) is the secret to building capital over time, for you and for the generations after you. (No matter how fun GameStop sounds.) Let's call a day *boring* if the high price that day is less than 3% different from the low price (that is, $\texttt{high} \leq 1.03 \cdot \texttt{low}$. I love boring days. I love going through multiple boring days in a row even more.

   Compute how happy I was last year. How many boring days were there? What was the length of the longest sequence of boring days in 2021? Print out the result in a reasonable sentence.

(As usual, if you do an extension problem, submit it by email directly to DLN.)

6. *(Entirely optional—for fun only.)* Extend your code from #3 to also figure out the highest high and lowest low of the stock price within each month, for January, February, . . ., December. What month has the highest volatility (the highest high minus the lowest low)?

   Think about how to solve this problem without copying and pasting the same code twelve different times! One option is to write a *function* that takes the month as its only argument, and then call that function twelve times. (Look up functions in Zelle.) Another option is to embed the code that you write in (another) loop that iterates over the twelve different months.

7. *(Entirelyer optionaler—for fun only.)* Figure out something (anything!) interesting in this data set using Python. Explain the interesting facts that you've discovered.