INSY 5336 001
Python Programming
Fall 2021
Final Term Project (100 points)
Due Date: December 6, 2020 11:59 pm CST (no exceptions)

The following guidelines should be followed and will be used to grade your project work:

- All code to be implemented and submitted as a jupyter notebook (.ipynb) file.
- This is an individual homework assignment, no group submissions will be accepted. If you discuss in groups, please write your code individually and submit.
- Sample runs shown in the question should be used as a guide for implementation. However extensive testing needs to be done on your code to deal with all test cases that might possibly be executed.
- The instructions for running of each cell and the expected results should be documented in the cell preceding the code using markdown language.
- Every code segment in the Jupyter notebook cells should be well documented with comments. Use # in the code to provide comments and they should explain the algorithm and what the code segment is doing.
- Error checking in your code is very important and differentiates a high quality programmer from a low quality one. Hence you should account for invalid user inputs, infinite loops, out of range results, etc. and resolve them by appropriate error messages. The homework will be graded for robustness of your code.
- Please read each assignment carefully. Note that you may need to test your code with example input files. I will be using my own test input file to test your code if needed. DO NOT hard code file names in your program except in cases where a specific filename is asked for.

This is a project to scrape data from the web and store the results in both a text file as well as the SQLite database.

1. (100 points) The WSJ's US Stock Market Movers website (https://www.wsj.com/market-data/stocks/us/movers) or (https://money.cnn.com/data/hotstocks/) tracks the most active stocks on a real time basis. You will first write Python scripts that collect the list of **Most Actives** tickers only from one of the above websites. Next, your programs should take these ticker symbols and build a comma separated text file (called stocks.txt) with data about each stock from the Google finance website:
2. https://www.google.com/finance/quote/AMD:NASDAQ which gives the quote for ticker symbol AMD as an example. The data to be collected from the Yahoo Finance site should include (Actual text as seen in the website is in brackets):
   OPEN price (Open)
   VOLUME (Avg. Volume)
   PE RATIO (PE Ratio (TTM))

   Note that the stock average volume is an INTEGER type but the quantity obtained from the website may have commas. Your program should remove these commas from the average volume quantity and store it as numbers.

   In addition to the stocks.txt file, the data should also be stored in an **SQLite** database called **StocksDatabase** in the directory that your Jupyter Notebook code will be executed from. The StocksDatabase should have a table called **StocksTable** that contains the following columns and types:

Ticker Symb	TEXT
OpenPrice	REAL
Volume	INTEGER
PERatio	REAL

Every execution of your program should create a stocks.txt and StocksDatabase.db file in the directory (delete any existing files that you will create) that your Python script is located and run. We will be testing your output by verifying the following:

1. The program gets the list of Most Actives ticker symbols from the https://money.cnn.com/data/hotstocks/ website and uses it to create the stocks.txt and SQLite database
2. Verify that the stocks.txt file is created with the Most Active ticker symbols and the data obtained from https://www.google.com/finance/ website in the comma separated text file format.
3. Verify that the StocksDatabase.db file is created and that the SQLite database contains the StocksTable with the appropriate columns and is populated with the data. SQLite Browser may be used to verify database content.