Computer Trading Strategies

Homework 2: Due 2/18/2025

Your assignment is to create a universe of stock data and then prepare an associated tutorial that specifies the procedures for creating it and the corresponding learning objectives, as well as identifying and resolving the set of issues that arise in completing the assignment. Please use the exact same procedures for preparing and submitting your homework as you did previously.

The stock universe itself is made up of the S&P 500 stocks. The universe of stock symbols is provided in the csv file 'SP Tickers.csv" that was provided with the dealing with data course module resources. You need to create an historical database, by day and for each stock, that contains the adjusted open, high, low, close, and volume. The date range should cover the last ten years as available in the Sharadar Equities Bundles available from Nasdaq Data Link. You should adjust the inclusion dates for each stock so that they conform to the time period that the stock was in the S&P 500. That is, only include pricing data for when a stock is in the S&P 500. Check the csv files "SP Additions.csv" and "SP Removals.csv" to obtain the valid date ranges. What checks are needed to help ensure the integrity/accuracy of the data? What error handling is needed? From this perspective, the starting point for the homework is the ending point of the class workshop, but you should include the workshop steps as part of the tutorial, but you should be brief about it if you are simply replicating what was covered in class. Of course, you could come up with a better way of doing it. In addition, you should incorporate the following data/functionality:

1. Add a column to your stock repository that specifies the financial sector (e.g., healthcare, energy, etc.) that the specified by a symbol operates. The sector for each company can found in the csv file called sectors that is provided with the dealing with data zip file.
2. For each stock and day in your universe of data, add in insider trading from the Sharadar equity bundle.

    a. Note that you can have multiple transactions for a given trading day.
    b. Your data set should only have one observation for each stock and day.

c. You will need to summarize the raw data in some logical manner that provides meaningful indicators/factors. Think in terms of # of insiders buying, how much purchased, # of insiders selling, how much sold, etc. Your tutorial should provide a rationale for how you summarize/transform the data.

3. For each stock and day, grab some data from any other data set in the Sharadar bundle of data (or elsewhere) and integrate it into your data universe. There is a lot of stuff in Sharadar alone, so think about what may be meaningful and helpful information in terms of making trades.

   a. One thought is to use the market capitalization (cap) data to calculate out the number of outstanding shares. Then you might use that to calculate out market cap for each day. This is simply one example of what you could consider doing.
      i. Be creative, but provide a justification/rationale for what you do, just as you do for the insider trading data.
      ii. Might require some consideration on how to deal with pricing adjustments going forward. You may need to keep track of the outstanding shares on daily basis.
   b. Separately, you may find it helpful to download all the fundamental data that you can get your hands on, and store it separately so that it will be available to you at a later date if you think of something useful you can do with it.