Hi Sunny –

I have a rather nuanced project that I’d like to get started on. I have three dataframe files that each contain data for about 100 stocks. They are all related and I will try and explain each of their respective purposes below.

The larger purpose of the entire project will be to use the dataframes to create a stock portfolio over the 10.5 year history of the data. The tricky part will be to incorporate the filtering rules that determine which stocks are a part of the portfolio for each day. There are approximately 100 tickers in the data files. However, only between 5-10 will be included in the portfolio on any given day. The price data for each ticker is included in: “df\_pct\_change”. While the filters for determining which stocks to include are derived from the remaining two dataframes that are also included.

Dataframes included:

1 - df\_pct\_change

This dataframe contains the daily percentage change of the closing price for each stock. The stock ticker is the column name, and the daily change from one day to the next is in each cell. The purpose of this dataframe will be to use this data for any of the tickers selected for a given day.

2 – df\_count

This dataframe indicates the number of days that a ticker has been “valid”. The values for each ticker will be either a running count or None. If the value is None, than this indicates that the ticker is not eligible to be included in the portfolio for that date. On the other hand, if the value were to be 20, then this would indicate that the ticker has been valid for 20 days and counting.

These numbers are relevant because they can be used to create filtering logic from. For example, if there are 10 stocks to choose from, we may want to pick the three that have been valid for the least amount of time (most recently became valid). Therefore, in this situation, those 10 stocks would be ranked in order from lowest to highest “count” value and the 3 with the lowest score would be chosen.

3 – df\_score

Each stock has a score given to it for all the days that the stock is valid for trading. I used historical performance data to calculate the score and the higher the score the better. The score can be used to create filtering logic similar to the count value. For example, we could choose to create a filter where out of ten valid and eligible stocks for trading, only 3 are chosen based on highest scores from this dataframe on the given day.

Specific goal

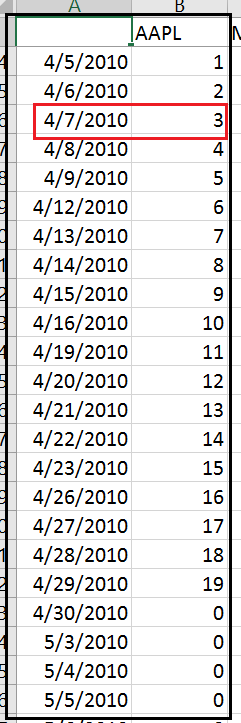
I have included some sample syntax that shows a portfolio being made by selecting the top tickers for each given day. By top tickers, I mean the ones on that day that are valid and have the highest score per the df\_score dataframe.

However, the problem here is that in a real life scenario, its not practical to choose a stock for only a single day… so we need to come up with some coding logic that will essentially buy and hold a stock in the portfolio until its no longer valid (the df\_count dataframe does a good job of representing if a stock is valid and for how long it has been valid for).

So, I think the goal for the first iteration of this project should be:

A script that will backtest through the entire date range of the dataframes (Jan 1, 2010 – Sep 14, 2020) apply the following “rules”:

1. A maximum amount of tickers that can be held on any given day (e.g., 10 tickers). This is a maximum… there could be potentially scenarios where this amount of tickers do not exist fyi, and this is ok.
2. A minimum score that the ticker must have on that day to be included in the portfolio (e.g., 7.5). The score values fluctuate for each ticker over time BUT do not fluctuate during the time when the ticker is valid for trading. Meaning, if there is a 20 day period where the ticker is valid, the score will not change during the 20 days.
3. This one I think will be the trickiest to implement… while a ticker is held in the portfolio, it can not be “sold” (and therefore it must remain in the portfolio) until it is no longer valid per the df\_count dataframe. Lets use the following as an example (please see image below)… let’s say AAPL is added to the portfolio on day 3 of it being valid. It can not be removed until after day 19 when it is not longer valid. It must be bought and “held” throughout the entire duration of it being a valid ticker in the portfolio. After AAPL is no longer valid, it is removed from the portfolio and another ticker that is valid can be added.



1. When there’s room in the portfolio for an additional ticker, the trading logic should isolate all eligible tickers that are valid and are not yet in the portfolio (ie ones that meet the minimum score threshold), and should add the ticker(s) to the portfolio that have the lowest “count” value (meaning, the tickers that have been valid for the fewest number of days)