CSCI 4022 Final Project Proposal

Andrew Pickner
10/25/21

I want to answer several related questions regarding the stocks of a given stock exchange to answer the question of what an optimal portfolio might be within that exchange. For starters, it would be nice to have a somewhat diverse portfolio including the top stocks from the top industries. Next, it would be better to have stocks in one's portfolio that are less risky or volatile. Finally, undervalued stocks are probably better than overvalued stocks unless that stock is consistently increasing in value by enough to disregard the overvaluation.

Although I took a peek at Google Finance, I'm not 100% if it is the service I'm going to use to collect all of my data. In fact, I may use multiple services and compile the data accordingly. Regardless of where I get my data, I'm going to get historical price data for all of the stocks within an exchange. Historical data will typically include the stock's ticker, date, prices at the high, low, open, close, and finally volume or the amount the stock was traded that day. I also want to include the sector that each stock belongs to, and either a trailing or guidance earnings per share. From here, I will be able to calculate the volatility scores, price to earnings ratios, and the daily percentage change in prices.

I'm probably going to need to talk to you about this further, but I will either use K-means clustering or the EM algorithm to cluster all of the stocks for each given day, and take the clustroid of all of these clusterings. I will most likely cluster based on the volatility score, P/E ratio, and potentially the daily percent change. Again, not 100% sure how I want to go about doing this, but I may also use the method I've described above for each individual sector, so I can compare it with the clustroid of the whole dataset to determine which industries are of highest importance.