1 Introduction

Why this is Necessary

At the core of finance lies the following question: Can anyone beat the market? To an economist, this is a question about the efficiency of the market. To most household investors, this is a question of whether $anyone$, even them, can make money in the stock market. This paper is a response to the latter interpretation of the question.

How to invest in the stock market is not something that is part of most people's educations, it deals with "real money", there are countless stocks to choose from, and household investors compete with hedge funds trading with complex algorithms. It is not hard to see why people would be a bit reluctant to risk their hard-earned money in an endeavor that looks perhaps less fruitful than playing poker against teams of professional players.

A fair amount of households do invest in the stock market, but through a secondary investor, like a mutual fund, in exchange for a commission. Yet historically, this is no better (most of the times worse) than if they had simply invested the money themselves into an index like the S&P 500. Historically, an index like the S&P 500 has been a very safe bet to go up in the long run, with average yearly gains much higher than those of bonds or the interest accrued from storing your savings in a bank. It doesn't seem to take very much skill to invest in a stock index, so it puzzles economists why more households aren't taking part in the stock market.

What this is

At the risk of naively oversimplifying things that we may not fully understand, in this paper we attempt to simply investing by using visual representations, rather than mathematical equations. Through this we hope to give household investors a way to explore the stock market beyond investing in a stock index, with any amount of money (as most comprehensive indices are thousands of dollars and not necessarily in the economic reach of most investors). It is fair to ask why visual representations are a "better" way than mathematical equations to represent the stock market. We don't claim it is "better." We simply claim that there is a lot of intuition that can be quickly and easily gained by studying the right graphs, and through this, investors can make investment decisions that roughly approximate those that maximize complicated formulas. Admittedly, this is a bold claim, but with that, let's begin.

2 The Big Picture

Where to Begin

Warren Buffett famously invests in companies he knows and trusts. Many investors will recommend that new investors stick to low-risk companies that are essential for every day life. While these are decent strategies, they are at least partially at fault for the mass under diversification amongst the portfolios of household investors [1] [2] [3] [4]. It is not enough to simply brainstorm a few household brands and invest in them. It is crucial that investors have a more concrete way of protecting themselves against the risk that they face by not diversifying enough.

So to begin let's list a couple of investment strategies: one is to invest in random companies. The other is to invest, as stated above, only in the household brands that first occur to an investor. As previously stated, the second strategy may fall prey to underdiversification (as an investor tends to think of similar companies if they were to brainstorm). The first strategy may provide good diversification (although there's no guarantee of this), but an investor would probably not be very confident in solely letting chance dictate whether they make or lose money. It can be naively argued that chance solely dictates whether one makes or loses money anyways. And this is true, but only to an extent.

Risk

We can call a stock's exposure to "luck" (basically how much the stock price tends to fluctuate around its "true" value, which we assume to be the mean) its risk. By our definition, this is exactly its variance. If a stock is more exposed to "luck" then when the stock is "lucky", the stock price is higher (compared to its mean) than a stock that is less exposed to "luck". This is one way of explain the link between average returns and risk.

This leads to an intuitive definition of diversification: a mix of high and low risk stocks, so that the overall portfolio has a certain level of risk and a corresponding expected average return. This is fair, but it isn't immediately clear why this is better than simply finding a single stock with a certain variance. The answer comes in the way we calculate the overall risk of the portfolio - it is not simply a sum of variances (risks).

Why is this the case? Consider buying and short selling the same stock. You are both betting its price will increase and decrease. Both buying a stock and short selling it carry some risk. But if you do both, the total risk is not the sum of both buying and of short selling the stock. Instead, if one action makes you money, the other loses the same amount, so if you sell and repay the stock at the same time, you are guaranteed to get your money back. Thus this action is pretty riskless. This is the idea that adding some stocks in the portfolio can actually decrease the overall risk of the portfolio.

[1] Home bias: *Investor Diversification and International Equity Markets* (French, Poterba, 1991).
[2] Local bias: *Local Does as Local Is : Information Content of the Geography of Individual Investors' Common Stock Investments* (Ikvovich, Weisbenner, 2005)
[3] Holding too few stocks: *Heterogeneity and Portfolio Choice : Theory and Evidence* (Curcuru, Heaton, Lucas, Moore 2004)
[4] Empirical results: *Down or Out : Assessing the Welfare Costs of Household Investment Mistakes* (Calvet, Campbell, Sodini 2007)

How does this work? The key is covariance. People like to think of each stock as a "random walk" in the short term [5], which effectively means that whether a stock goes up or down tomorrow depends on nature flipping a coin. But not all the coin flips are independent. One coin flip can affect another coin flip and make one result more likely or unlikely. This is covariance. In the extreme case, one stock goes up if the result of a single "coin flip" is heads while another stock goes down with this results (this is like the scenario of buying and short selling the same stock). The idea is that we can choose stocks, so that not matter what the result of the "coin flips" are, we have stocks that go both up and down in varying amounts, giving us a positive return on average, no matter what the coin flips result in.

Covariance

From our explanation of covariance, we should be able to easily see covariance by looking at how stocks together (or in the opposite direction).

Let's try it out by looking at how 53 stocks performed over the course of a year (from April 4, 2019 to April 4, 2020):

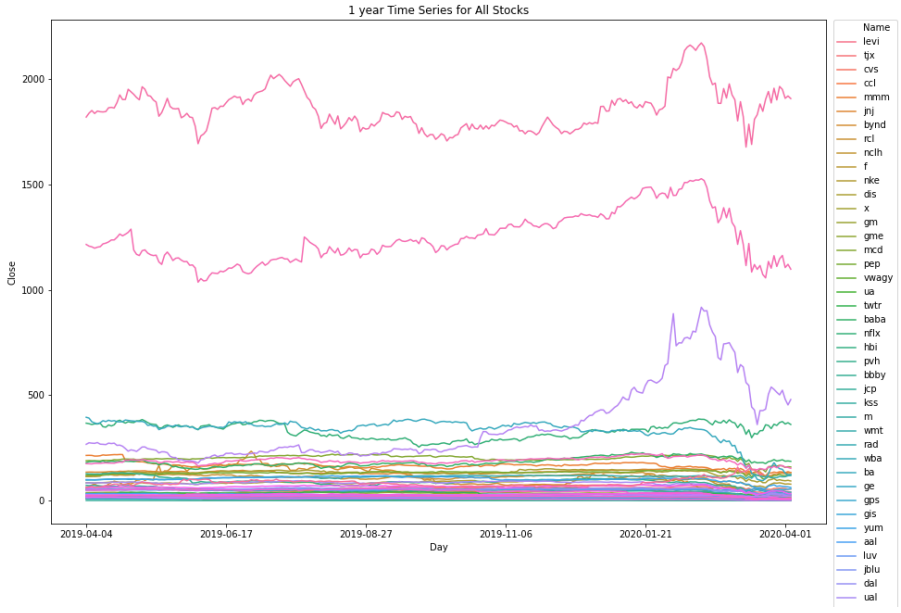

Figure 1: What do we see? Not much. Even with only 53 stocks (a tiny fraction of the stock market), graphing them is overwhelming and not particularly helpful.

One of the most apparent problems from the previous graph is that some stocks are a lot more expensive than other stocks, so large fluctuations in one stock (ex. from $5 to $10) are dwarfed in magnitude by stocks

[5] *A Random Walk Down Wall Street* (Malkiel 1973)

that are order of magnitude more expensive. We can attempt to alleviate this by instead looking at the day-to-day percentage change in a stock's closing price relative to the previous day's price. Basically if a stock went from $\$5$ to $\$10$ from one day to the next, then the percent change is $\frac{10-5}{5} \times 100 = 100\%$. Let's see the percent change from day to day of the same 53 stocks over the same period.

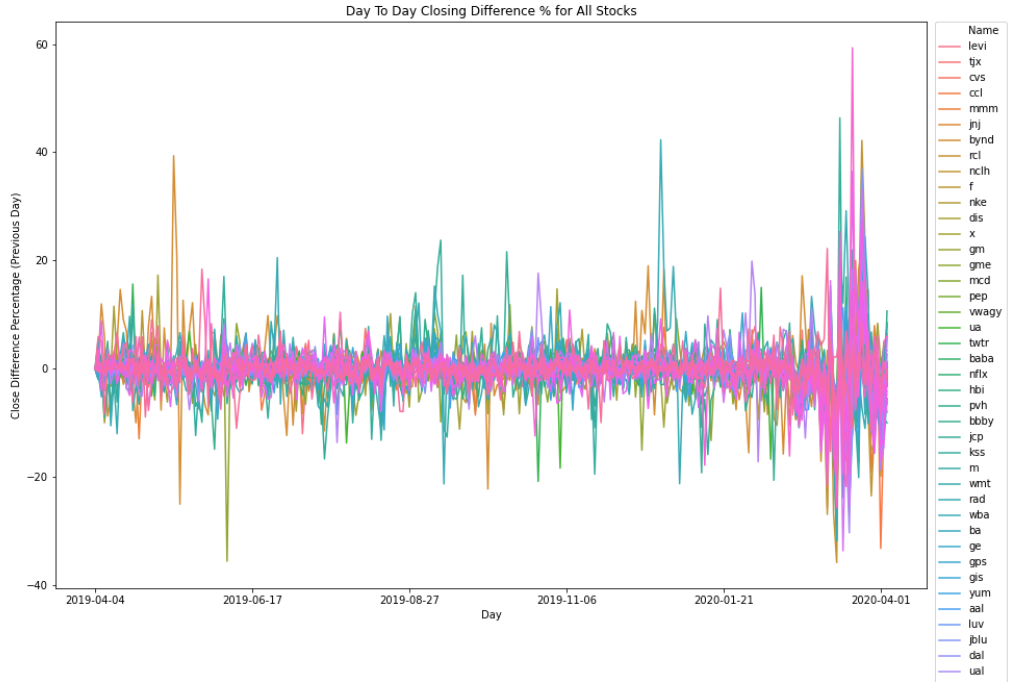

Figure 2: It is a cool graph, but it is nearly impossible to tell what is going on.

From just the two figures above, it is clear that the stock market can get very overwhelming very quickly. It is no wonder many household investors stick to 10 or less stocks, which are much easier to visualize and keep track of. But this does not mean that visual representations of the stock market are useless. It means that the visual representations we are looking at are not the right ones.

It is common to see visual representations of the stock market of either a single stock over time or of a stock index over time. These are easy to interpret, but give zero insight into covariance between stocks in a portfolio, which is key when choosing which stocks to add to your portfolio. There is a simple fix, however: rather than comparing individual stocks, compare individual industries (specifically indices of each industry constructed by adding a bunch of stocks of companies in that industry).

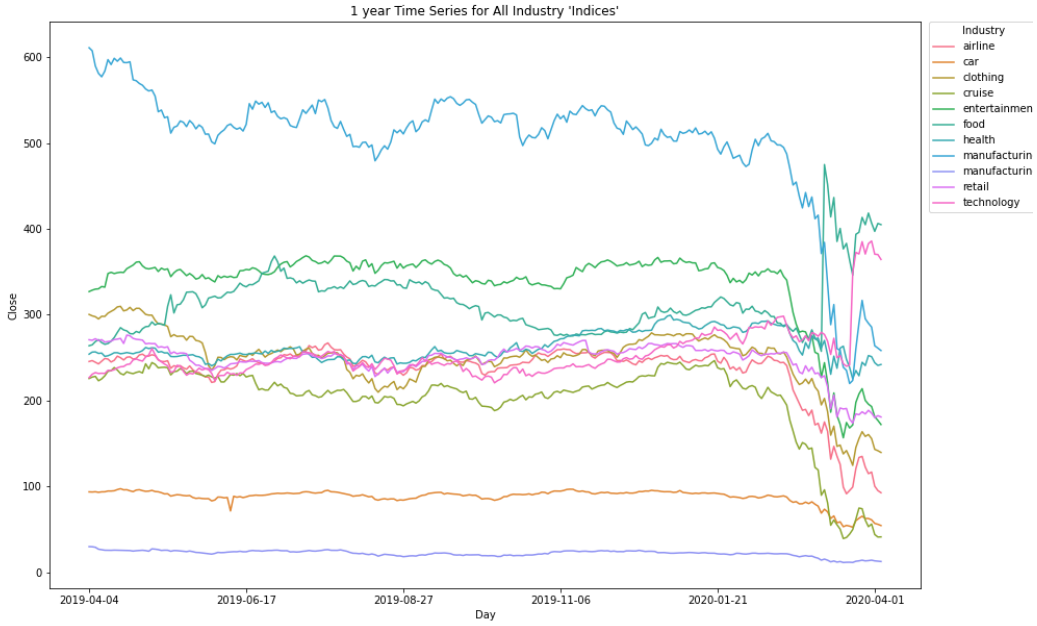

Figure 3

This data is much easier to look at. We can see that in general, every industry took a big hit in March of 2020 due to the economic impact of Covid-19. But before that, most industries were trending upward. However, the technology and entertainment industries actually went up while other industries were plummeting due to Covid-19. Perhaps this is unique to the Covid-19 response featuring people stuck at home, not going to physical stores, but increasingly consuming content online. Either way, this is clear covariance in opposite directions between technology stocks and all stocks that depend on physical products. This means that you can hedge against the effects of another shut down by investing in technology and entertainment stocks.

If we want to focus now on the variance, or risk, within an industry, it is helpful to look at histograms that show us how often an industry stock index was at a certain price (if the price stayed the same throughout the entire year we expect a single peak while we expect a rectangle representing a uniform distribution if the price was unpredictable throughout the year).

In Figure 4 we see the distributions of the industry index price. We can think of these as showing the distribution of your holdings, if you had purchased one of each stock in each industry at the beginning of the time period (April 4th 2019). From this, we can see the distribution of those combined prices over this past year. Histograms like these are a great way to see the variance over a specific time length, but we can't see when certain prices occurred. But looking at these graphs, we see that most industries have one dominant peak, indicating that the prices in general are steady. The width of the peak is related to the variance, or the risk factor. We see that the wider it is, the more likely the price is to go down, or up. We see that with quite a few, such as airline, car, cruise, entertainment and retail, there is a tail on the lefthand side, and sometimes even a hump. This represents an overall decrease in stock prices in that industry and for a few industries such as the cruises and airlines, we can assume that this is an artifact of the COVID-19. Now, there are some very interesting industries that don't have one clear hump. We see in food and technology that there are in fact two main peaks, indicating that combined, the stocks in those industries have dramatically changed from one stock price to another. With food, this is the result of one company, Beyond Meat. Now, lets compare these individual industry level indices to a total index of all of the companies we had included and see if the distribution can tell us anything.

As we see from Figure 5, you would need a decent amount of money to be able to invest in all of the companies listed (this is the main problem with stock indices that are "safe bets" and have good diversification, which motivates looking for a way to more strategically diversify with more limited capital). Besides that, we do thankfully see one main peak indicating that the overall the together, the prices are generally going to hover around one area and doesn't have a tendency towards another extremely different price as we saw with some of the industries. However, we do see that there is still a fairly wide base, ranging from about $5,500 to $8,500 and to most people $3000 is a lot of money. Let's try to quantify this by looking at its standard deviation, which is 462.97. The standard deviations of each of the above industries is

```
airline : 35.85
cars: 9.00
clothing: 35.01
cruise: 46.73
entertainment: 45.69
food: 190.44
health: 16.16
manufacturing: 72.32
manufacturing2: 3.54
retail: 20.00
technology: 164.75
technology2: 193.20
technology3: 173.97
```

And the sum of the variance of each industry is 1006.68. This is much higher than the standard deviation we reported for the total "index". Why is this? If you have a background in probability, you would know that we can't simply sum up the various individual standard deviations unless they are independent. As mentioned previously, when the industries aren't independent (and there no reason to assume they would be, as it is generally agreed upon that all stocks have a common component), we have to take into account the covariance between the various industries. Thus, the calculation for the variance (and thus standard deviation and risk) of the total index involves calculating the covariance between the industries. As we mentioned previously, the time series graph of industries implies that while some industries were massively negatively impacted by Covid-19, like cruises and airlines, some industries went up, like technology and entertainment. So the overall portfolio benefited from the diversity in industries and was not as severely impacted by Covid-19 as a portfolio consisting of solely cruise stocks or airline stocks. This is the benefit of diversification.

3 Looking Closer

Which Stocks to Pick

We've seen that there is a quantifiable benefit to diversifying across industries due to some industries being able to handle certain economic events (like the shut down due to Covid-19) better, thus softening the blow when other industries are hit hard. This is a good argument for investing in a stock index with stocks from various industries. But we have yet to provide a good alternative for those that either can't afford to invest in a large index, like the S&P 500, or want to try their luck hand picking companies to invest in (keeping in mind the added risk each stock adds to the overall portfolio).

But we can do just this if we use the previously seen visual representations to pick industries that seem to co-move minimally with each other (or even better, co-move in opposite directions) and then use more industry-specific visual representations to pick stocks within an industry that co-move minimally with each other (or again, even better, co-move in opposite directions). Effectively, the logic for picking individual stocks within an industry to invest in is exactly the same as the logic to choose industries to invest in. The goal is to ultimately reduce the overall risk of the portfolio by taking advantage of diversification.

It may not be entirely clear why reducing the overall risk of the portfolio is the goal, if after all, we measure "success" in the stock market by realized gains, not the variance of the portfolio. The thing is, we don't know what the realized gains will be before they are, well, realized. All we can do is estimate based on previous data. In general, we would only invest in a stock if we estimate its price to increase, so we take it as a given that all the stocks we consider are expected to increase in price if we add them to our portfolio (since it is up to the individual investor to decide which stocks they expect to rise in price; there is no defined rule for this as for any stock transaction between two investors, the investors must disagree on whether the stock will rise or fall in the future). If we expect a positive return from every stock we add to our portfolio, then we are essentially saying that we expect "luck" to be "on our side"; we expect "good luck" in the long run average. But we would like a bit more assurance that we will get good result, or at least that if we have "bad luck" in the short run, the results will not be disastrous. That is, we expect good things are more likely, but nothing is guaranteed, so we would like to strategically minimize the risk of bad things happening. This is why we diversify.

In fact, the relationship between risk and reward is even stronger than this. If we wanted to minimize risk we could simply not invest in the stock market at all. It turns out that the "rational" view of finance dictates that the expected (or average) return of a stock is solely determined by the risk involved in the stock (or its variance)[6]. Since we tend to be risk-averse, we must be compensated for undertaking more risk. This is effectively manifested by lower demand for riskier stocks (because of the risk aversion) and thus a lower "market" price. This lower price now effectively increases the expected return later (think of it like this: if you buy a lottery ticket, your expected return is higher if you pay $\$0.01$ for the ticket than if you pay $\$100$ for the same ticket and chances of winning the same price).

This means that we want assets that are individually risky if we want a higher expected return as the expected return of a portfolio *is* additive. That is, we add up the expected return of each stock in the

[6]CAPM: $Capital Asset Prices : A Theory of Market Equilibrium under conditions of Risk (Sharpe 1964)$

portfolio. This expected return is the "fair" return corresponding to also adding the variance of each stock in the portfolio. But as we saw before, due to diversification and covariance, the actual risk involved in a diverse portfolio is less than we are compensated for. This is what we want. This is effectively what differentiates the stock market from gambling in a casino. The odds don't have to be against you. With more informed investing, you can expect (on average) to get significantly higher returns than you get by investing in bonds, keeping your money in a bank, or even giving your money to a secondary investor. Moreover, it does not have to take solving complex differential equations or even calculating the fair value according to the CAPM to choose a non-naive portfolio that takes into account risk.

We leave you with industry-specific plots. We will not go in depth analyzing them, but rather leave it as an exercise to the reader. These plots capture a snapshot of the market in the middle of an economic shut-down imposed due to Covid-19, so they should be of particular historical interest. This also means that we don't expect any strategies made from these plots to be necessarily applicable for anything other than strategies to protect against future global crises. However, the increased variance during this time should make co-movements between stocks abundantly clear (see the cruise stocks for nearly perfect co-movement).

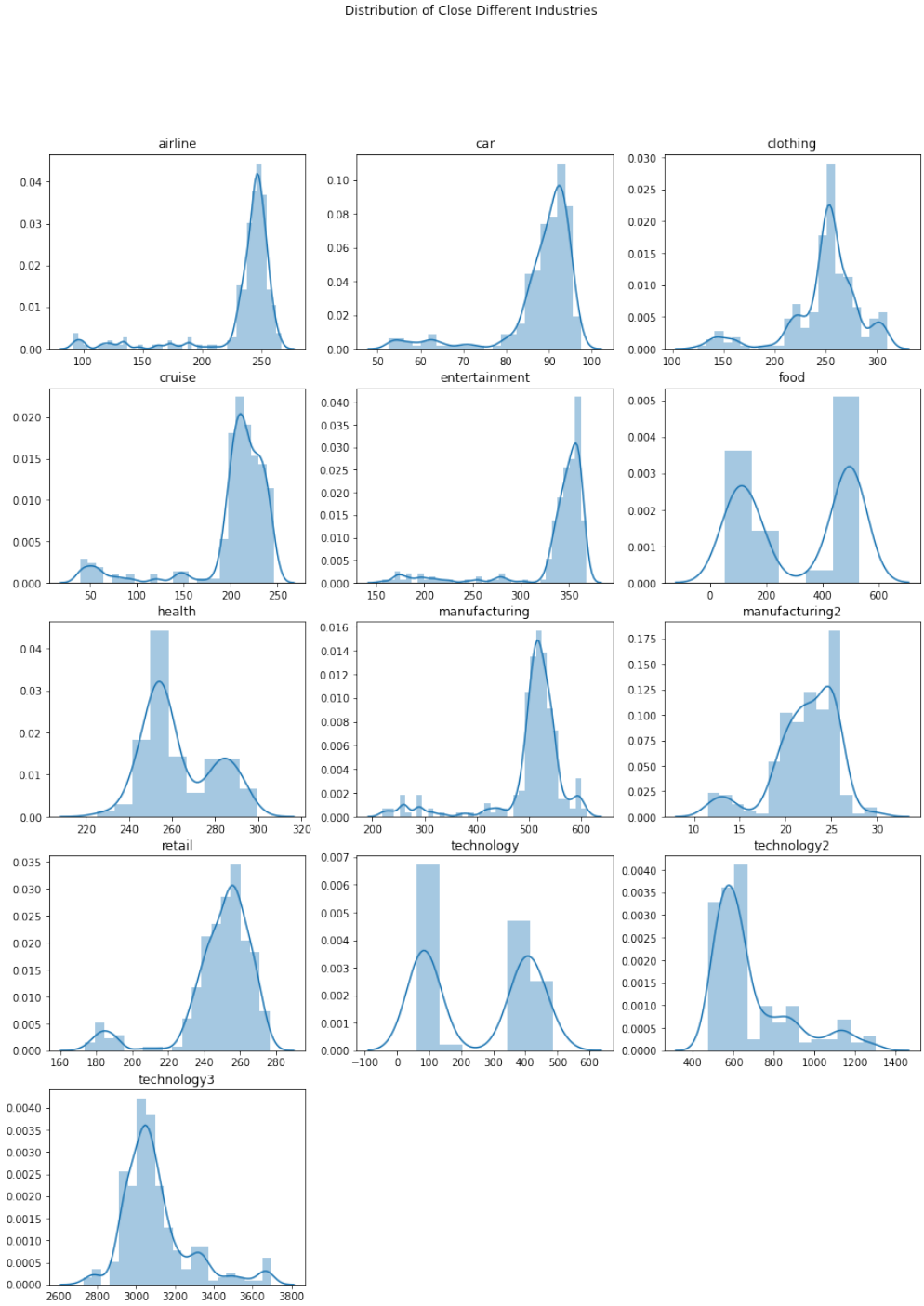

Figure 4: Distributions of the closing price for various industry "indices" created from adding the prices of various individual stocks together.

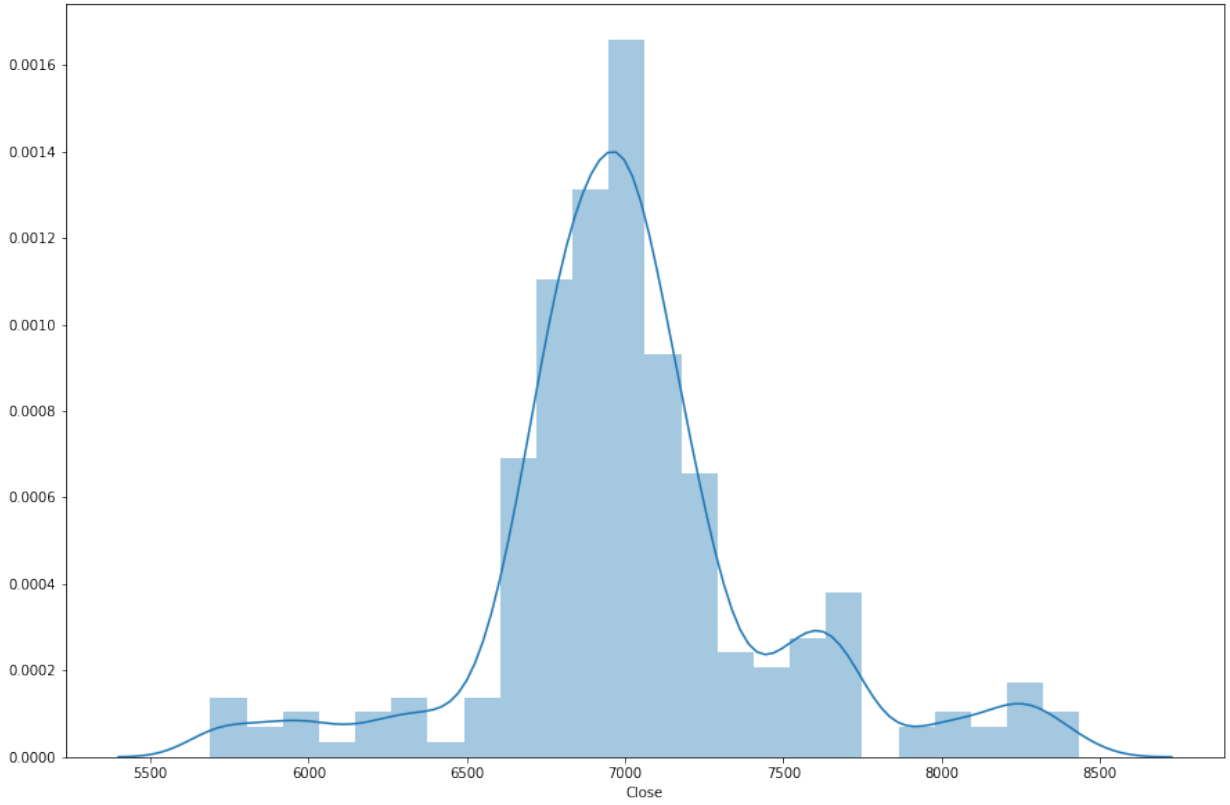
0.0016
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
5500
6000
6500
7000
7500
8000
8500
Close

Figure 5

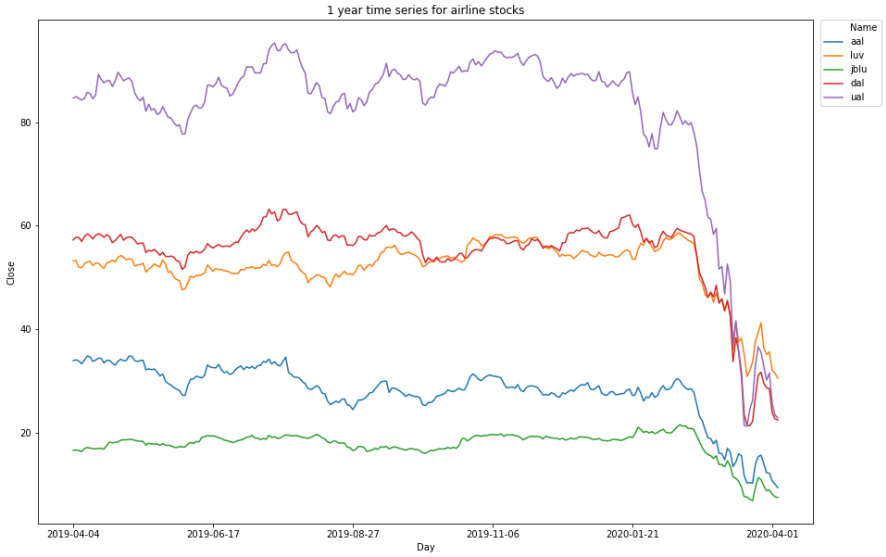

Figure 6

Figure 7

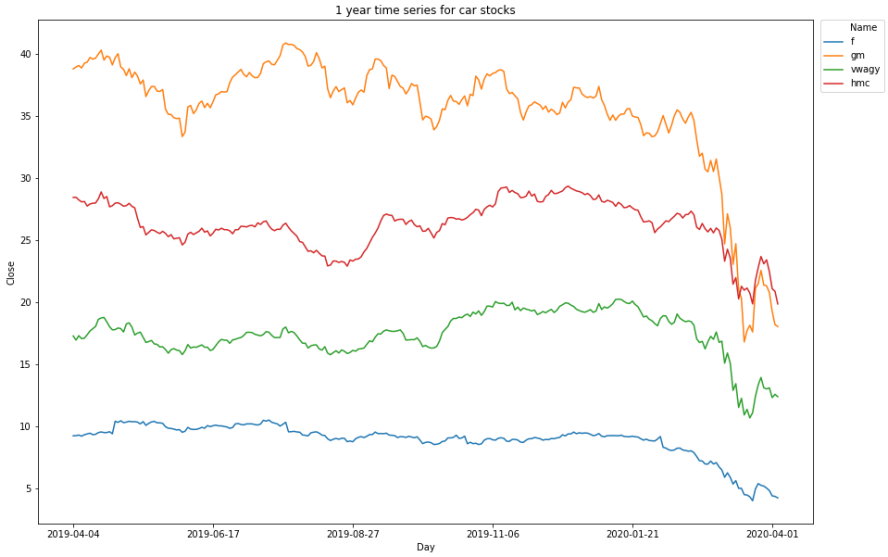

Figure 8

Figure 9

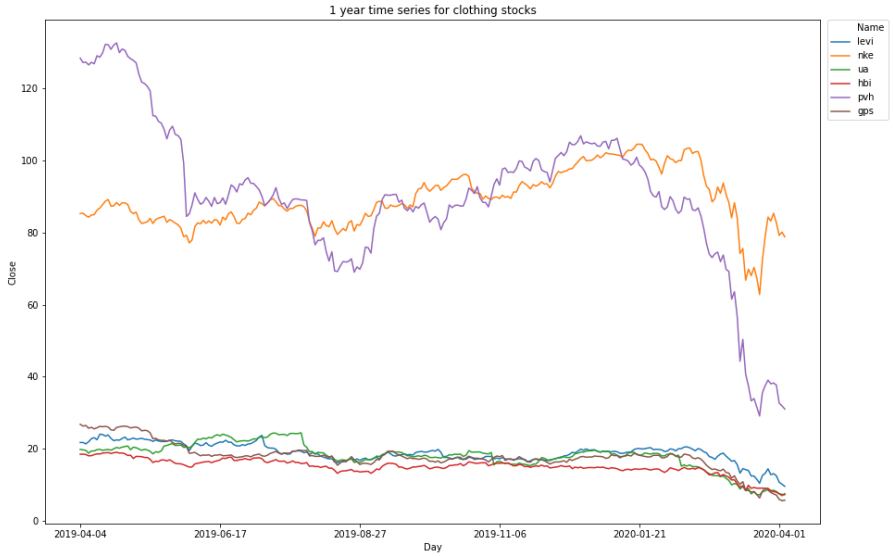

Figure 10

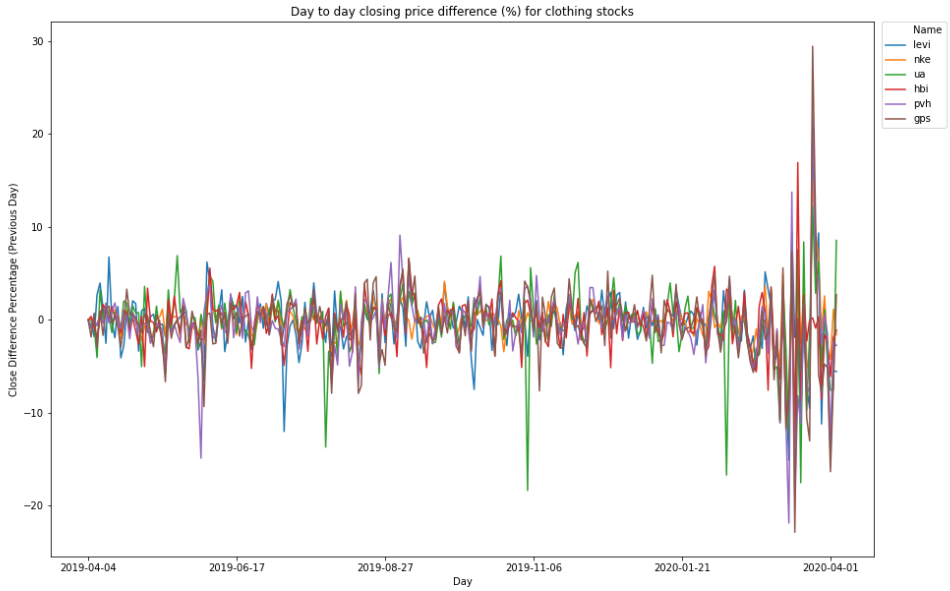

Figure 11

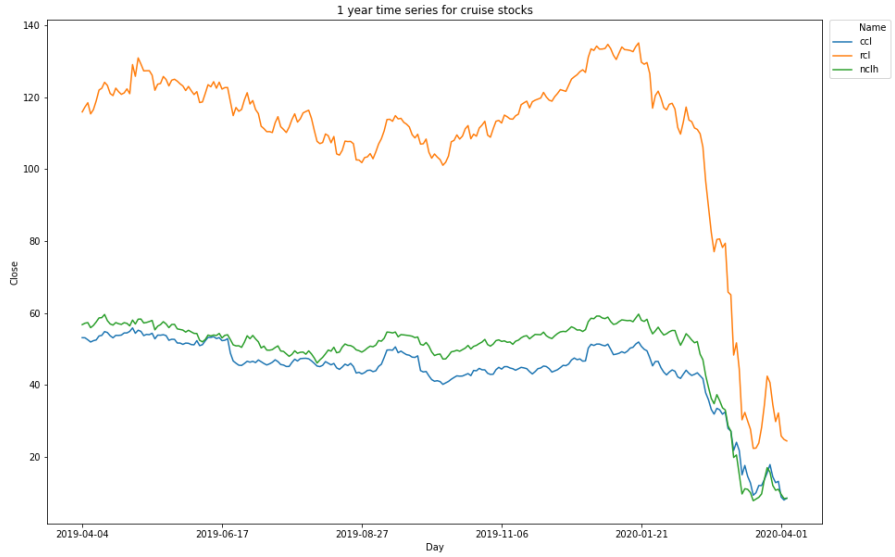

Figure 12

Figure 13

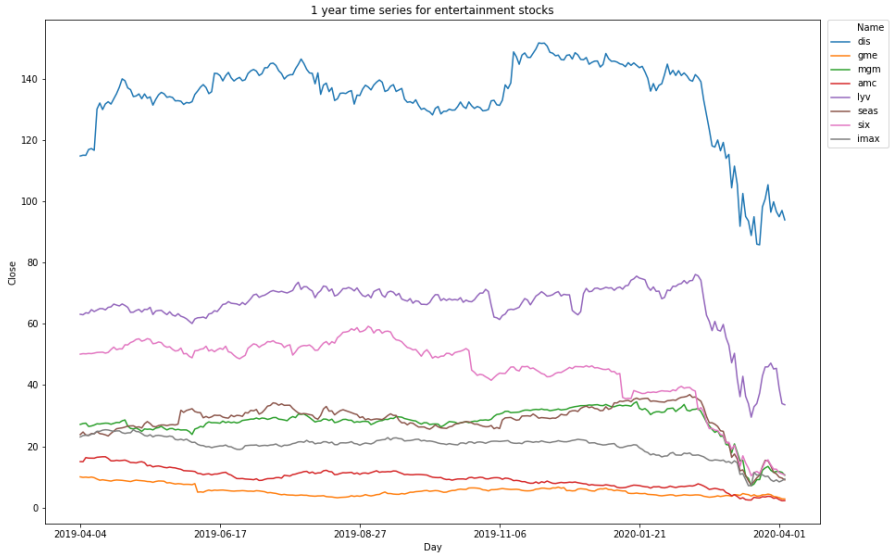

Figure 14

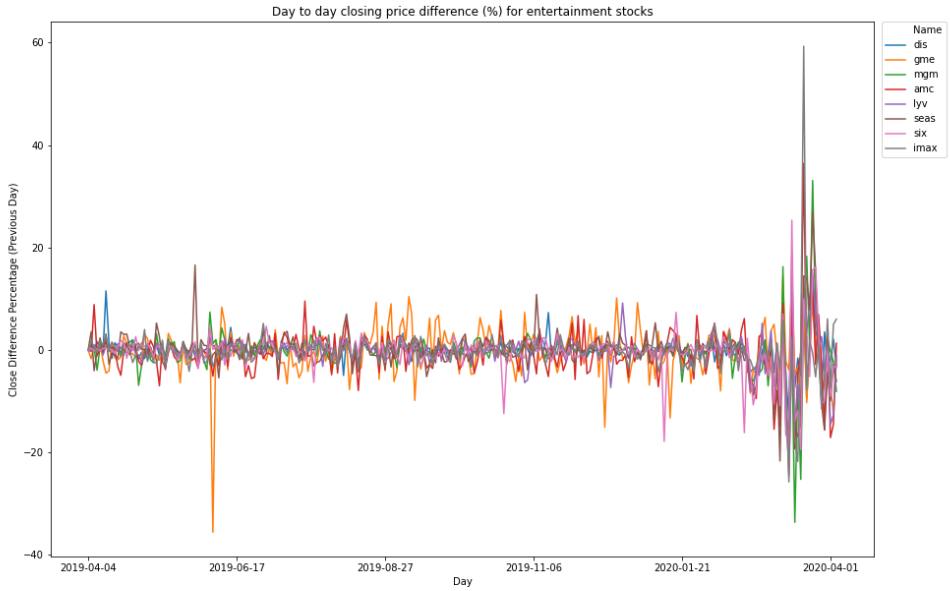

Figure 15

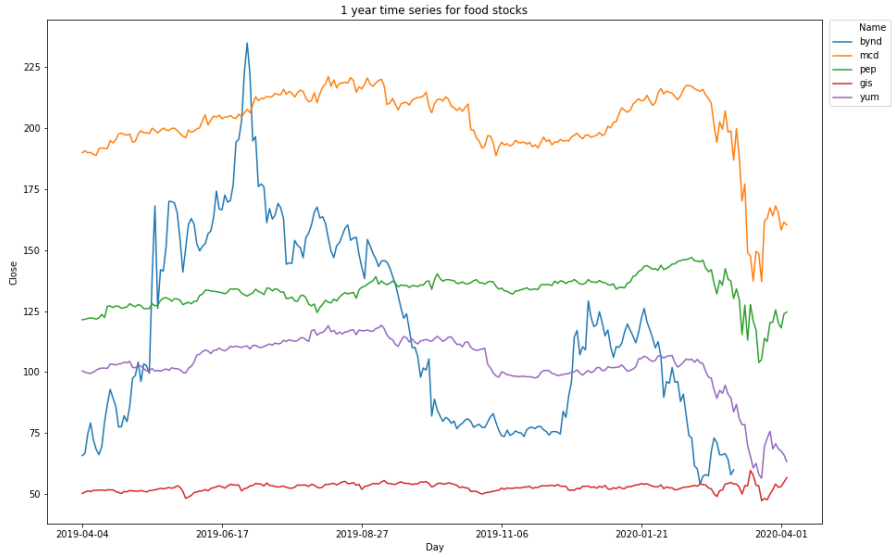

Figure 16

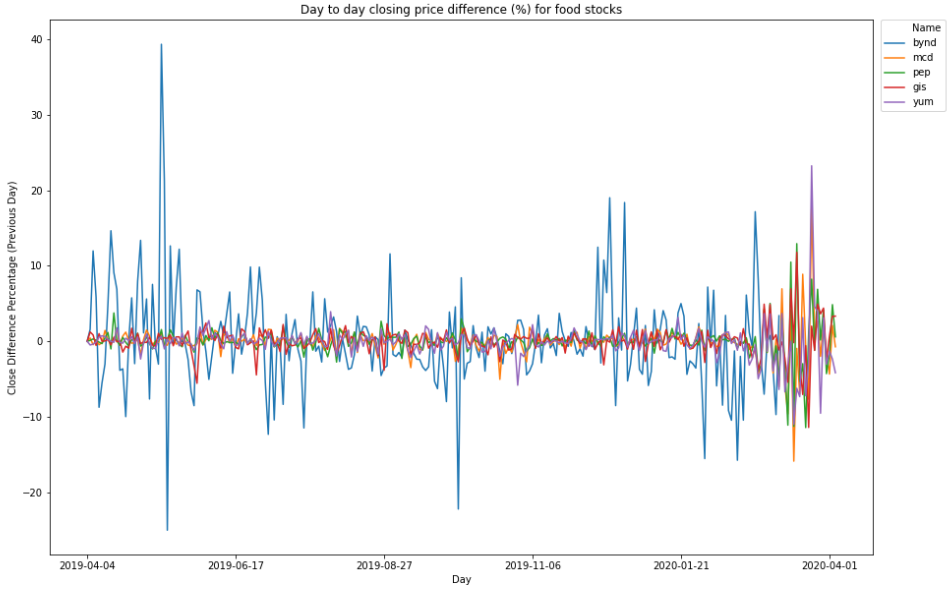

Figure 17

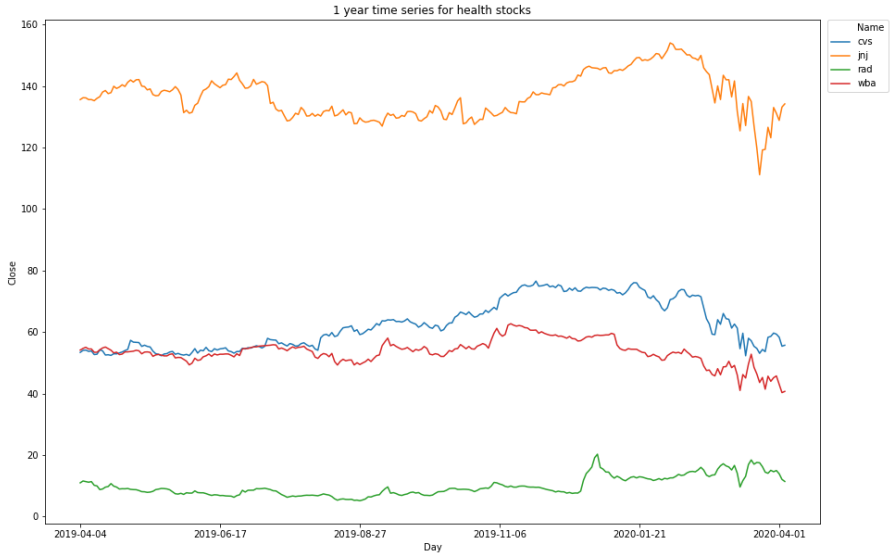

Figure 18

Figure 19

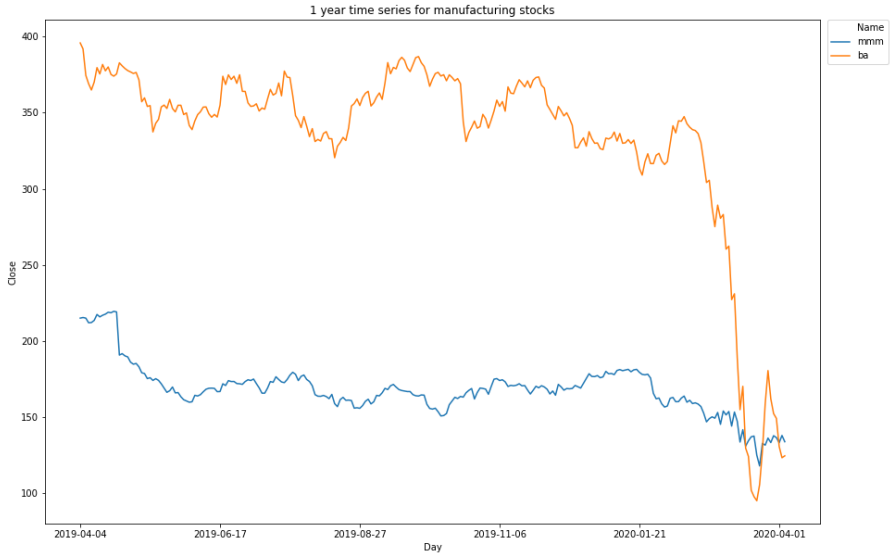

Figure 20

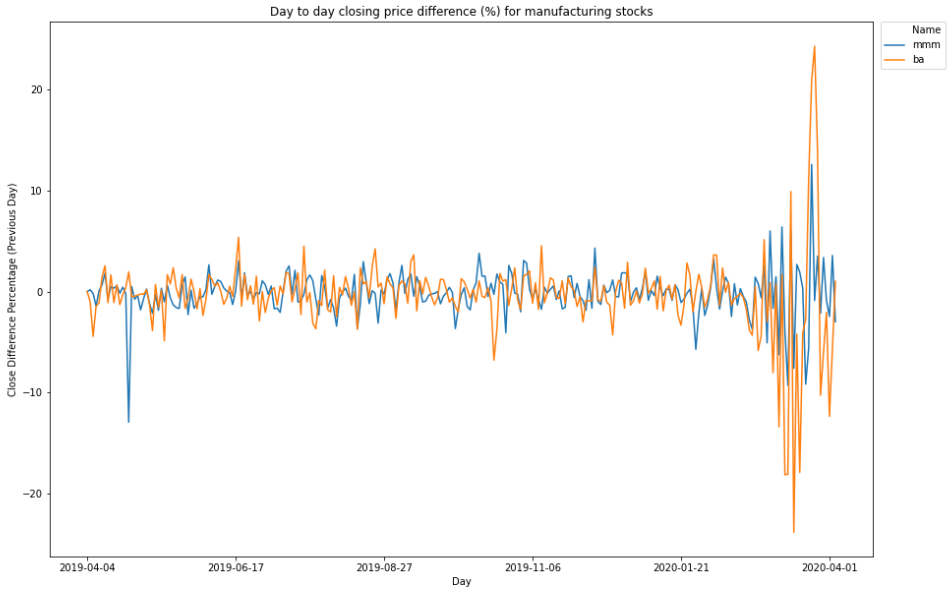

Figure 21

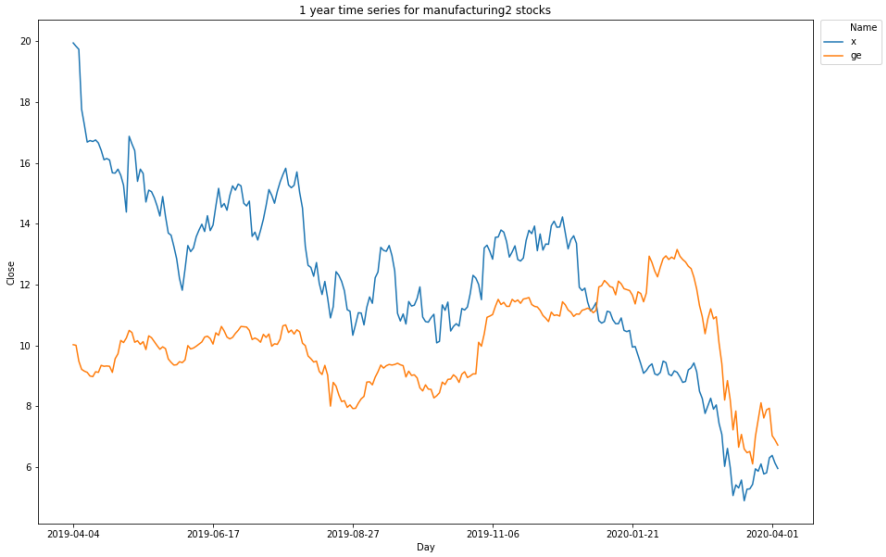

Figure 22

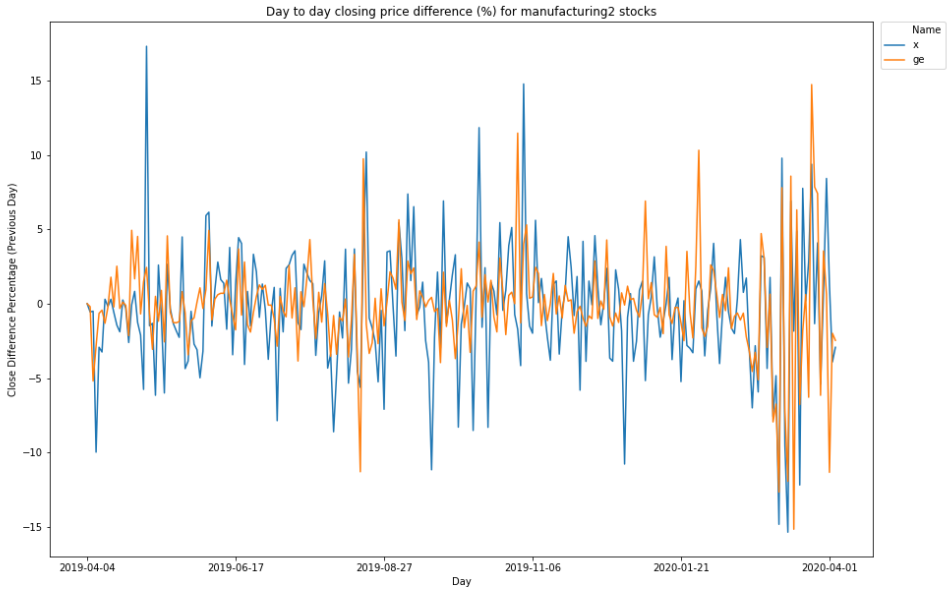

Figure 23

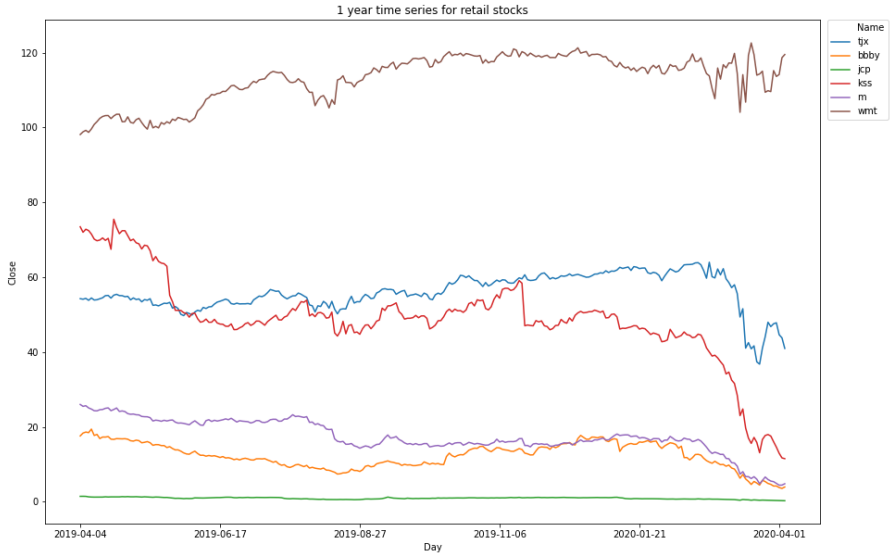

Figure 24

Figure 25

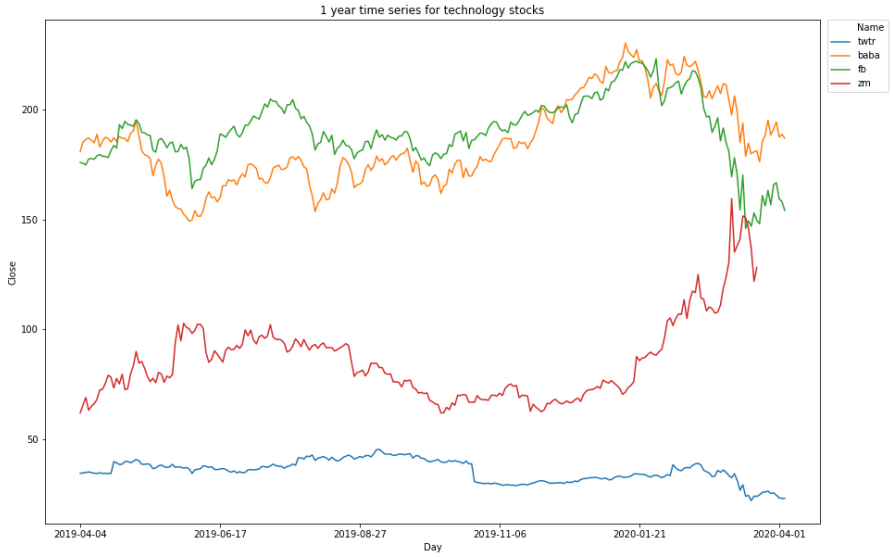

Figure 26

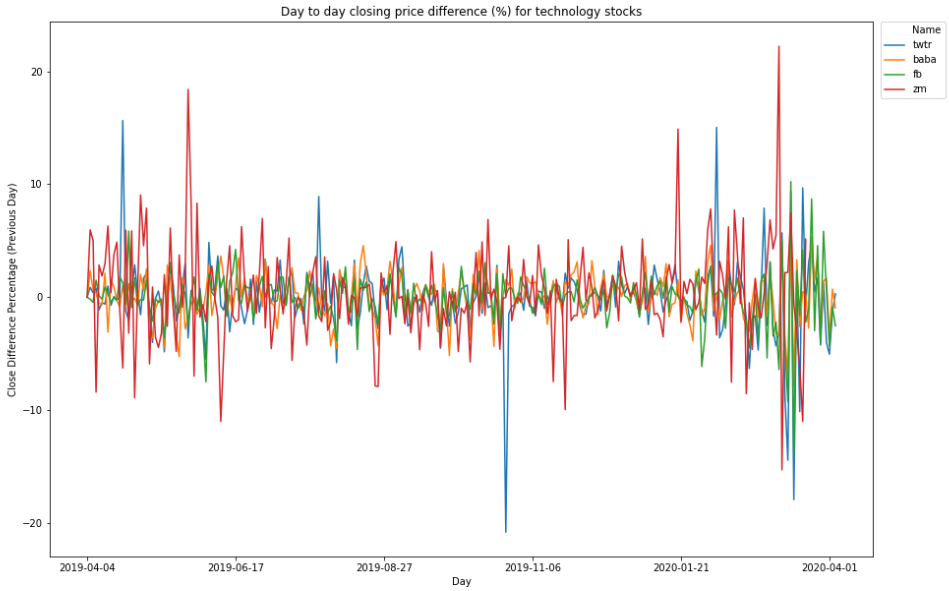

Figure 27

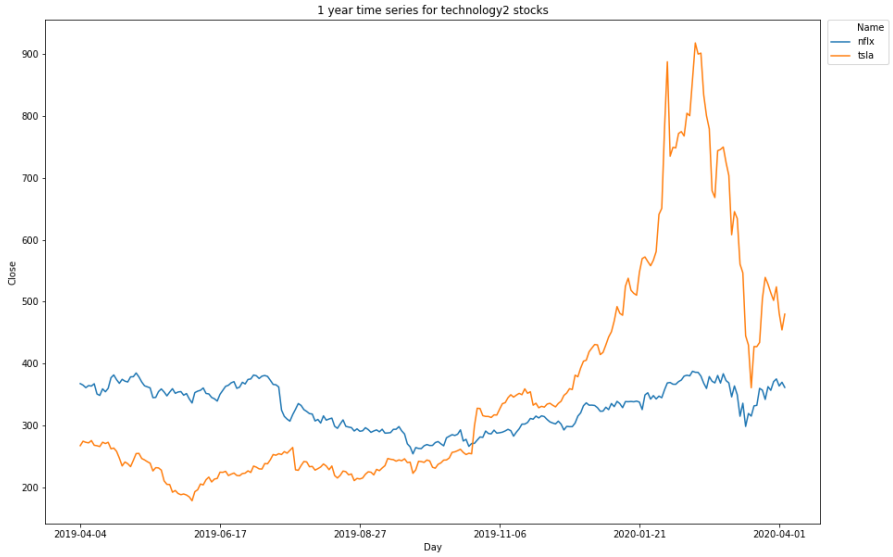

Figure 28

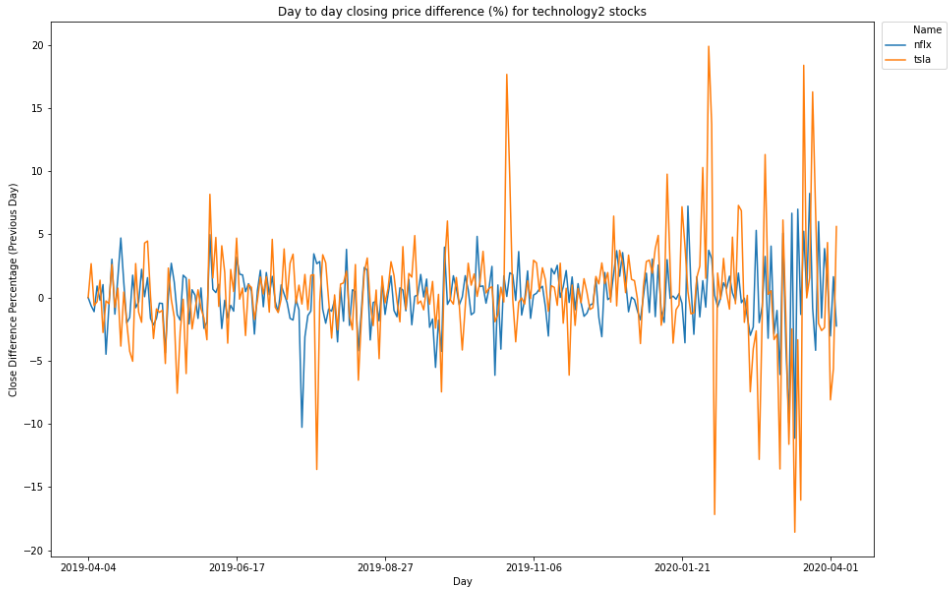

Figure 29

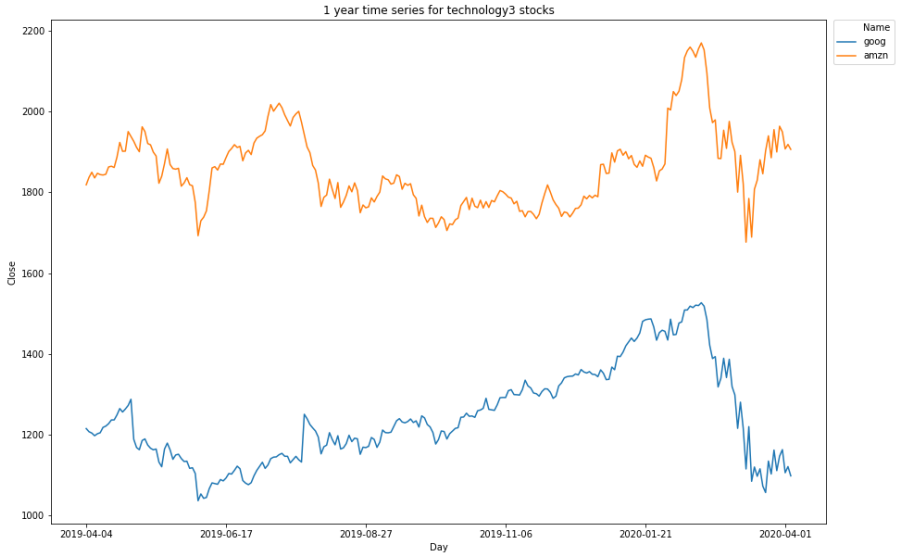

Figure 30

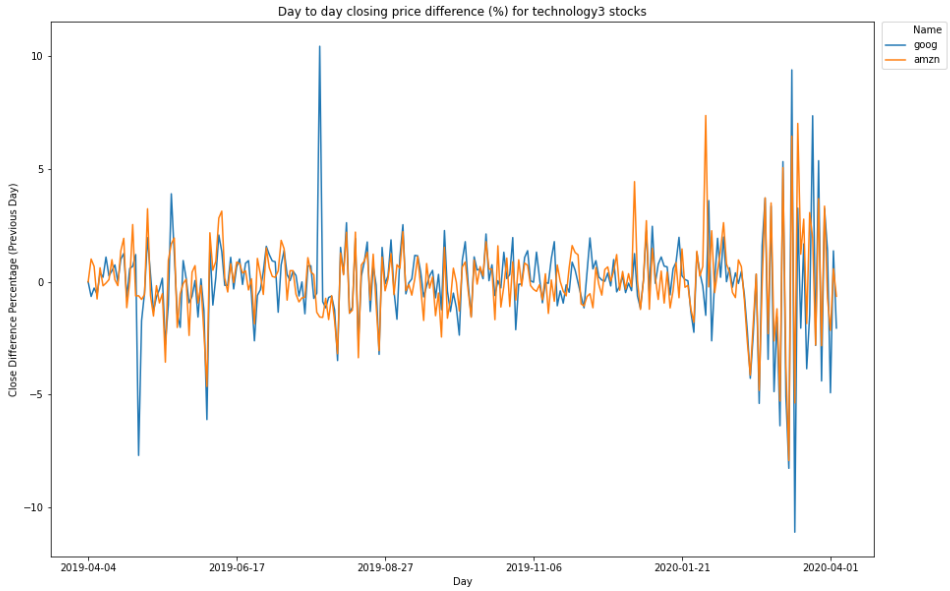

Figure 31