Stock market investing is complicated, especially when you don't know what, or when, to invest. Our goal is to analyze historical stock market data and draw insightful conclusions into which types of stocks & sectors one should invest their money, as well as analyze any potential indicators to predict stock growth (or recession). Stock valuation discrepancy exists in the market due to different market capitalization and industries. Our team is to dive into the question by examining various factors and provide analysis as holistic as possible.

- What is the stock market general trend?
  In order to make Jupyter notebook run faster and more efficiently, we limited our scope to examining the general stock market from 2000 to 2018. And here we are only examining the trend of trading volume of different industries within this period. In technical analysis, volume measures the number of a stock's shares that are traded on the stock exchange in a day or a period of time. Volume is important because it confirms trend direction.

  The interplay between stock trading volume and stock price usually can tell some insightful stories:

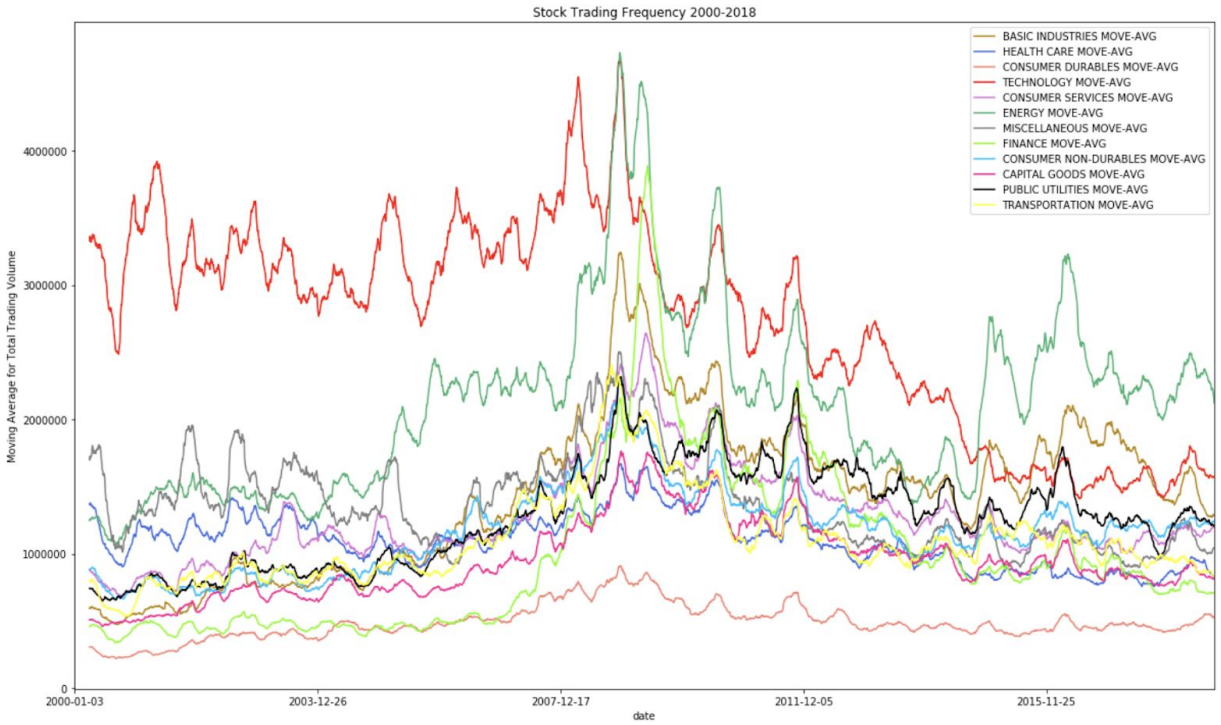
  When both stock price and volume increase, it shows the stock's uptrend. Some positive earnings news and people's expectation of price increase make people desire the stock more.

  When the stock price increases and volume decreases, it indicates trader's indecision or indifference to buy the stock. It might be due to people's uncertainty of the company's potential and future performance. The trend is unstable.

  When both stock price decreases and volume increases, it usually shows a downtrend of the stock. Maybe both earning news and people's expectations are both very negative.

  When both the stock price and volume decrease, it indicates trader's indecision to buy the stock. The trend is unstable and could change.

From the graph "**General Stock Trading Frequency 2000-2018**."



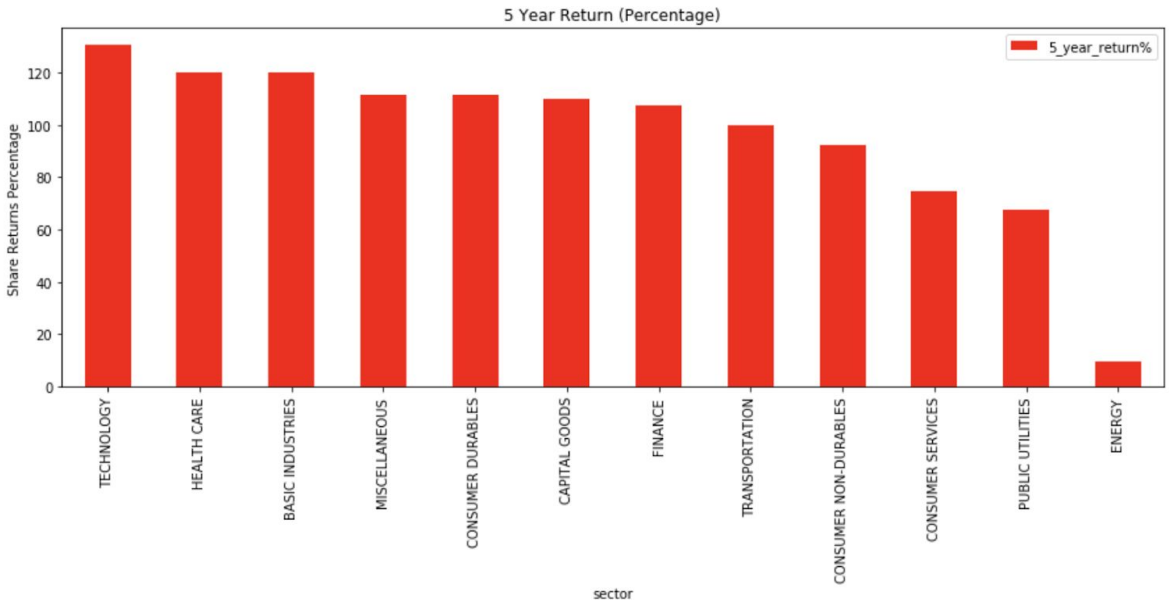
We can see technology sector's trading volume was stable and significantly higher than any other sector around early 2000. the trading volume in technology sector keep declining from 2000 to 2004. It matches the early 2000 dotcom boom and consequently dotcom burst. However, the trend recovered until 2007 and then kept declining again even more severely since 2008 financial crisis.

When at peak, Finance and Energy sector's reaching volume reached as high as Technology sector's trading volume before declining since 2008 financial crisis. However, energy sector had a significant recover around 2015. This might due to two massive and permanent shifts in energy around 2015. The first one was OPEC on mission to destroy long-term competition and preserve market share. The second was the fortification of a move toward alternative energy as 200 nations agree to sign a U.N agreement to avoid dangerous climate change.

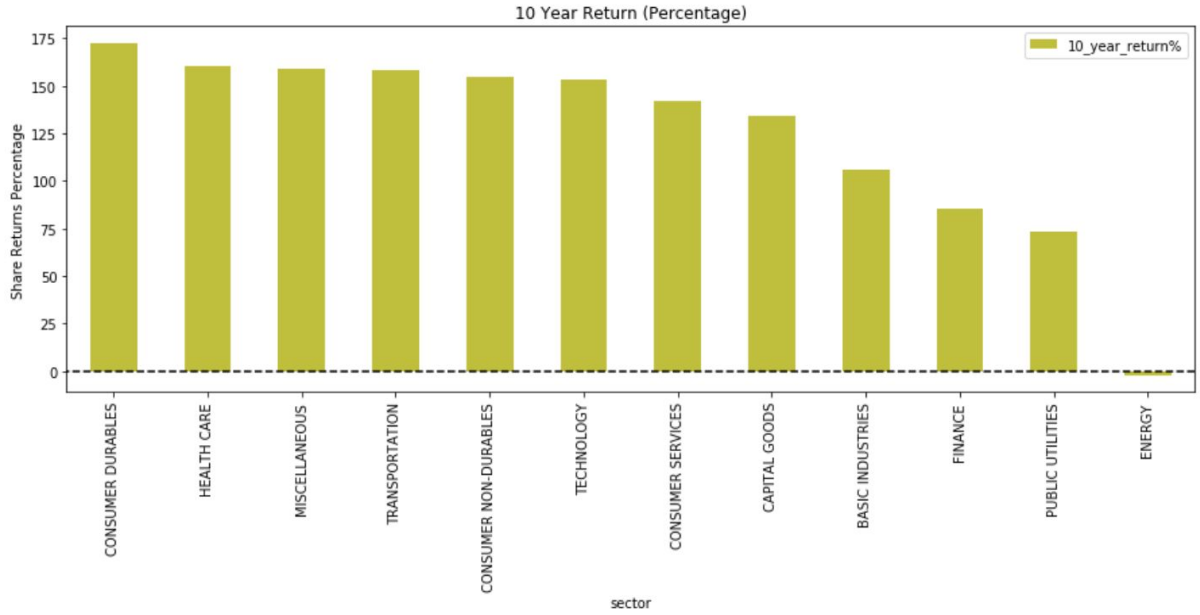
On the other hand, consumer durable sector's trading volume has always been stable since 2000. Consumer goods are usually everyday necessities. People will always need those goods no matter how the economic situation is. Those companies' stocks are usually low risk, low return. You cannot count on this sector and make a fortune, so the trading volume is also relatively low all the time. People can always add some of the stocks from this sector to their portfolio for hedging purposes.

- What industry sees the best return? (5-year, 10-year,15-year)
  Next we are examining the five-year, ten-year and fifteen-year returns of different sectors.

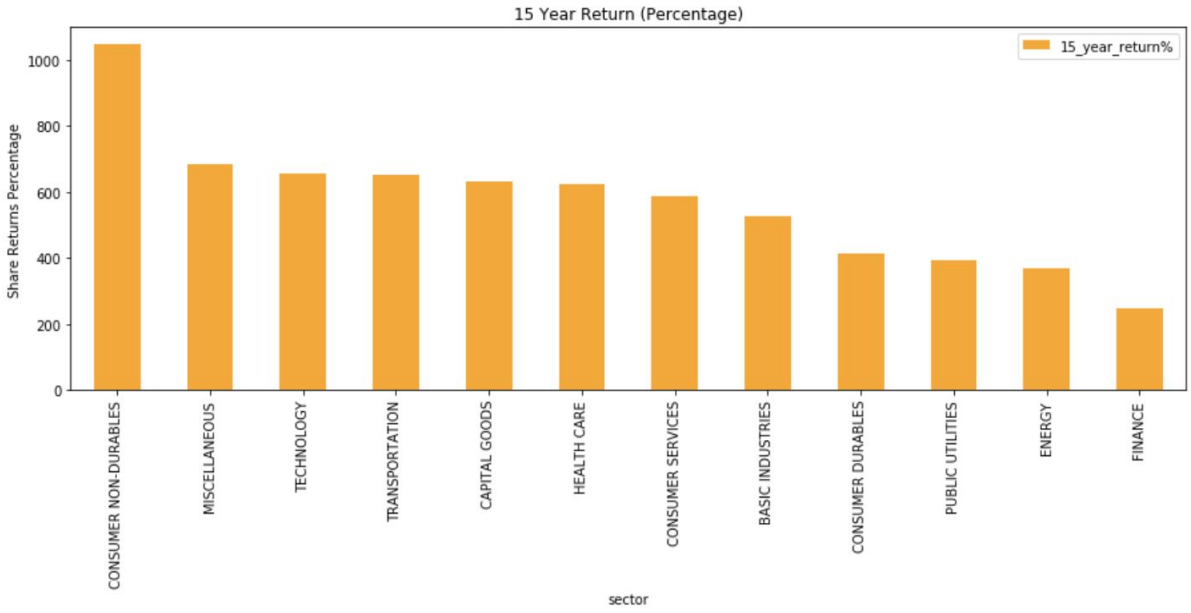
  First of all, we analyzed the five-year return from 2013 to 2018 and visualized the data. In the chart “5 year Return (Percentage), we can clearly see that sectors like technology, healthcare and finance have very strong average return. Public utilities sector also shows a decent average return of approximately 70%. Public utility companies’ stock prices are usually not high and don’t have significant increases, so this average five-year return is somewhat surprising. The reason might be due to the stable dividends. Sometimes, the dividends of those utility companies can reach to 5% to 10%. It’s a good strategy to incorporate some utility stocks into a portfolio to hedge risk and have stable passive income.



Next, we examined the ten-year average return for different sector within period 2008 to 2018. Sectors like technology, transportation and healthcare also shows similar average return as in five-year chart. Finance sector's return is not as high as in five-year chart. After 2008 financial crisis, finance sector needed time to recover slowly.



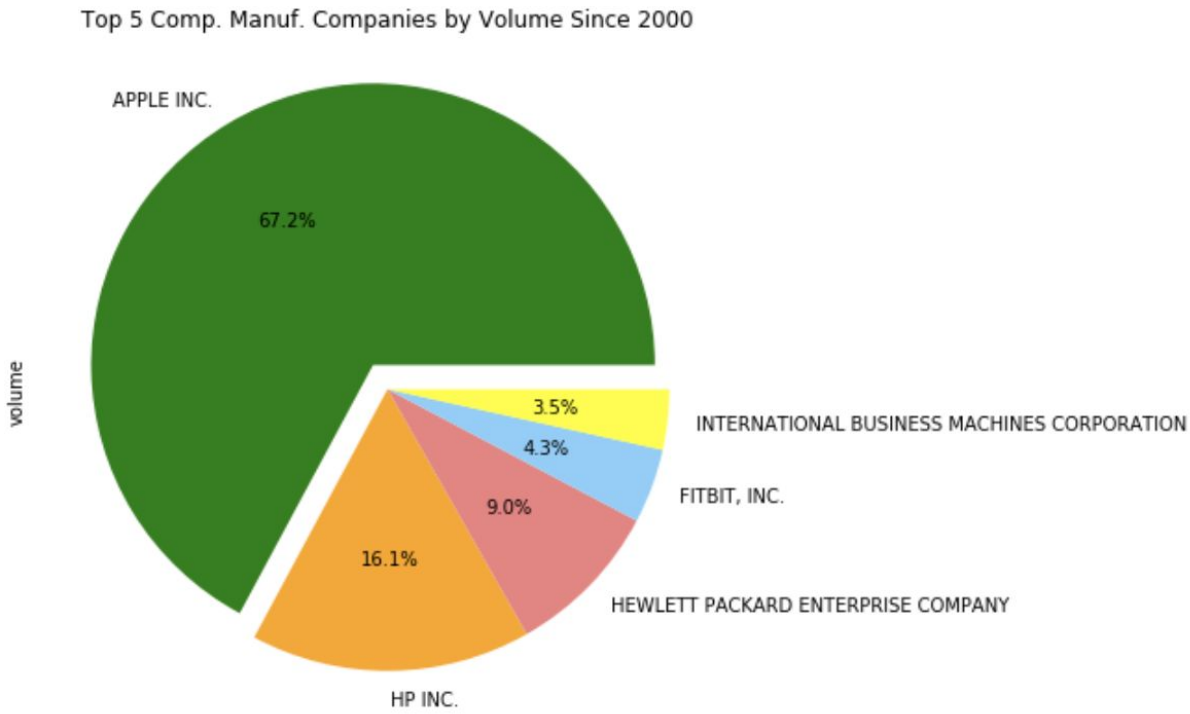
Lastly, we analyzed the fifteen-year average returns for different sector within period 2003-2018. Consumer non-durable sector shows a very strong and significant higher average return than all the other sectors. Consumer non-durable goods are usually products which has expected life-span of less than three years. For instance, food, tobacco, clothing, personal care products. The demand for those goods is usually much less sensitive to the economic situation than for other goods. Consumer non-durables sector is comprised of companies with more consistent operating histories and more consistent records of dividend increases. Moreover, consumer non-durables are generally lower-growth businesses with above-average dividend yields. No matter dotcom or 2008 financial crisis, it does not make this sector as volatile as other sectors such as finance or technology.



- Compare the individual stocks, which stock is the most popular since 2000?

  Since we are in the Bay Area, next we want to investigate further into technology sector. After analyzing the csv file using Pandas, we found out that computer manufacturing industry under technology sector has the highest average trading volume, followed by computer communication equipment and semiconductors industry. Furthermore, computer manufacturing industry also displays the strongest ten-year return from 2008 to 2018. Given that computer manufacturing has both highest trading volume and 10-year return, we want to take a detailed look of the companies within this industry.

  The top companies within computer manufacturing industry with highest trading volume include Apple, HP, Fitbit, Dell, TeraData. We plotted a pie chart showing the trading volume of the top five companies:



From the pie chart, we can see that Apple has overwhelming advantage over other competitors within this industry. Apple has the highest trading volume which is even significantly higher than the second place HP. From the data frame below, Apple also shows a very high ten-year price change.

Higher volume for a stock is an indicator of higher liquidity. High volume stocks are usually good stocks that have low liquidity risk. More people are trading these stocks because more people want them in their portfolio.

The above information demonstrates that Apple has both high trading volume and high 10-year return. It means that over the 10 year period, Apple's earning is generally satisfactory and people also have an overall optimistic attitude towards the company's future.

Limitation of Average Annual Return:

When you had an investment worth $1million at the beginning of 2017, by end of 2017, this investment was worth $3million, with a 200% average annual return rate. In 2018, the company didn't do well, stock price dropped, your investment lost 50% ended up with $1.5 million. What's the return rate on your investment over this period? (200%-50%)/2=75% ? Average Annual Return does not work here, so we need to use CAGR.

What is CAGR?

CAGR means Compound Annual Growth Rate. It is a mathematical formula that shows a smoothed rate of return.

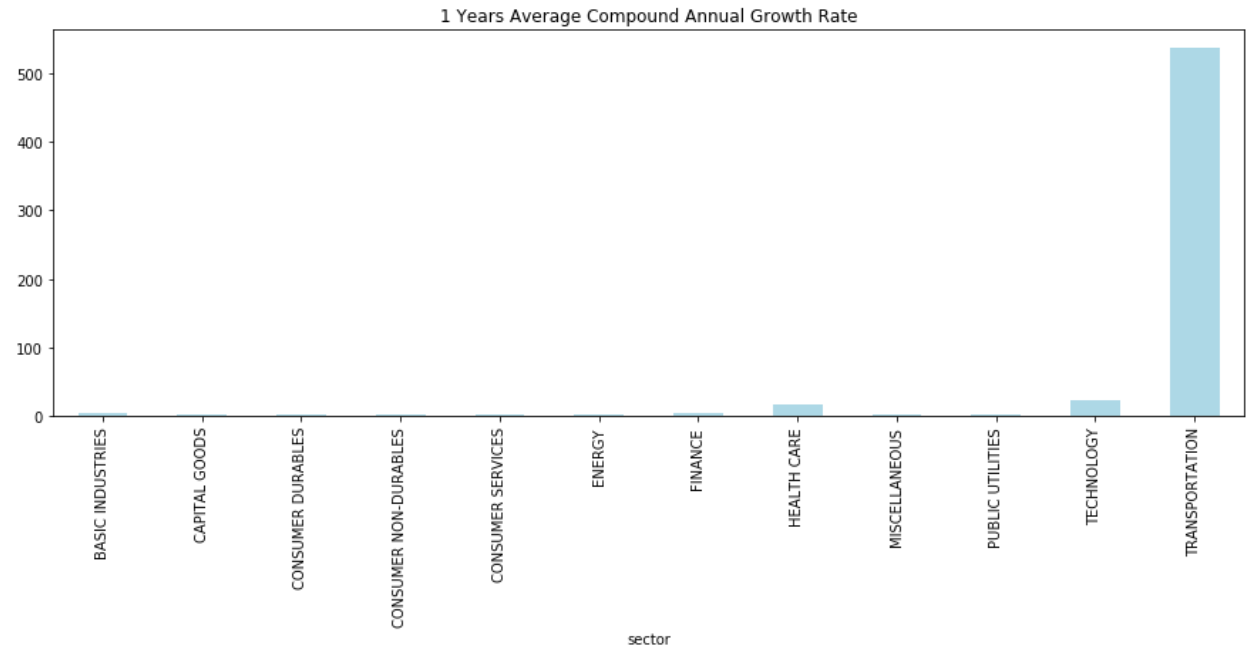
CAGR = [(EV / BV) ^ 1/n] - 1

EV=ending value of an investment; BV=beginning value of an investment;
n = number of years

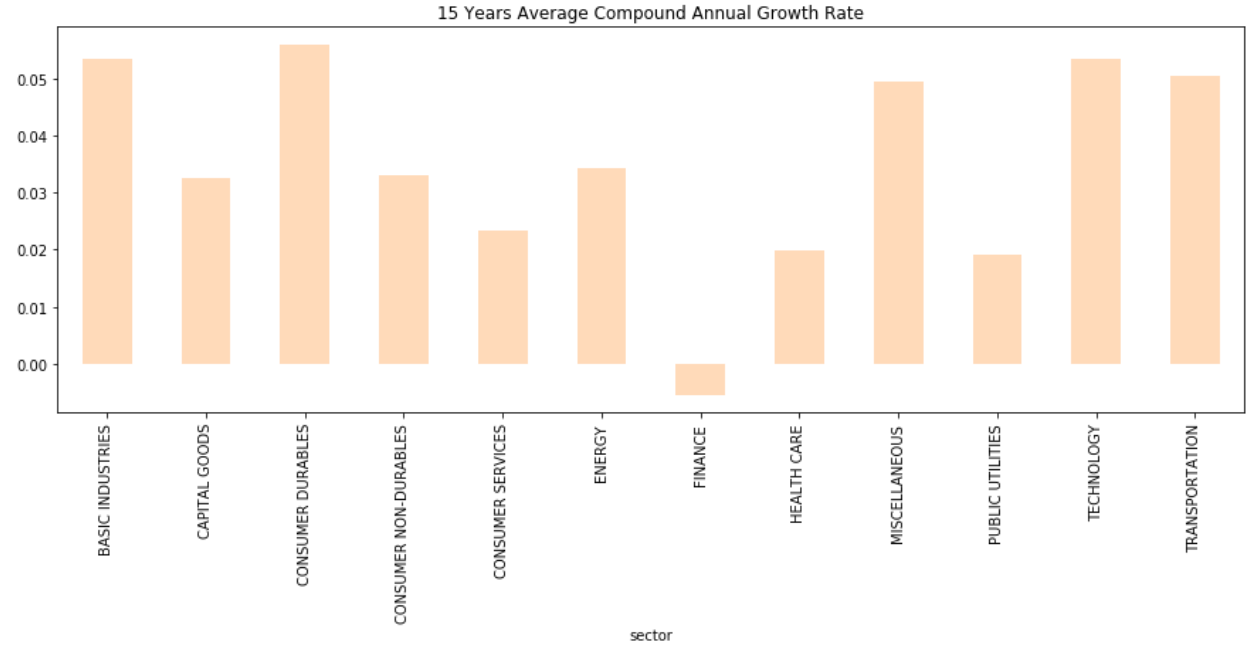
CAGR isn't a true return rate, but rather a representational figure. It is essentially a number that describes the rate at which an investment would have grown if it had grown the same rate every year and the profits were reinvested at the end of each year .It is used to measure the value of an investment and its growth over a given period of time to estimate the total return on a particular investment. While the annual rate of return gives the amount of return every year, CAGR gives the return over the entire period of the investment. CAGR can be used to smooth returns so that they may be more easily understood when compared to alternative investments. As a result, investors are better able to evaluate the returns on different investments before making a choice on which works best for them. CAGR is the best formula for evaluating how different investments have performed over time. It helps fix the limitations of the arithmetic average return. Investors can compare the CAGR to evaluate how well one stock performed against other stocks in a peer group.

Insignificance of one year CAGR:

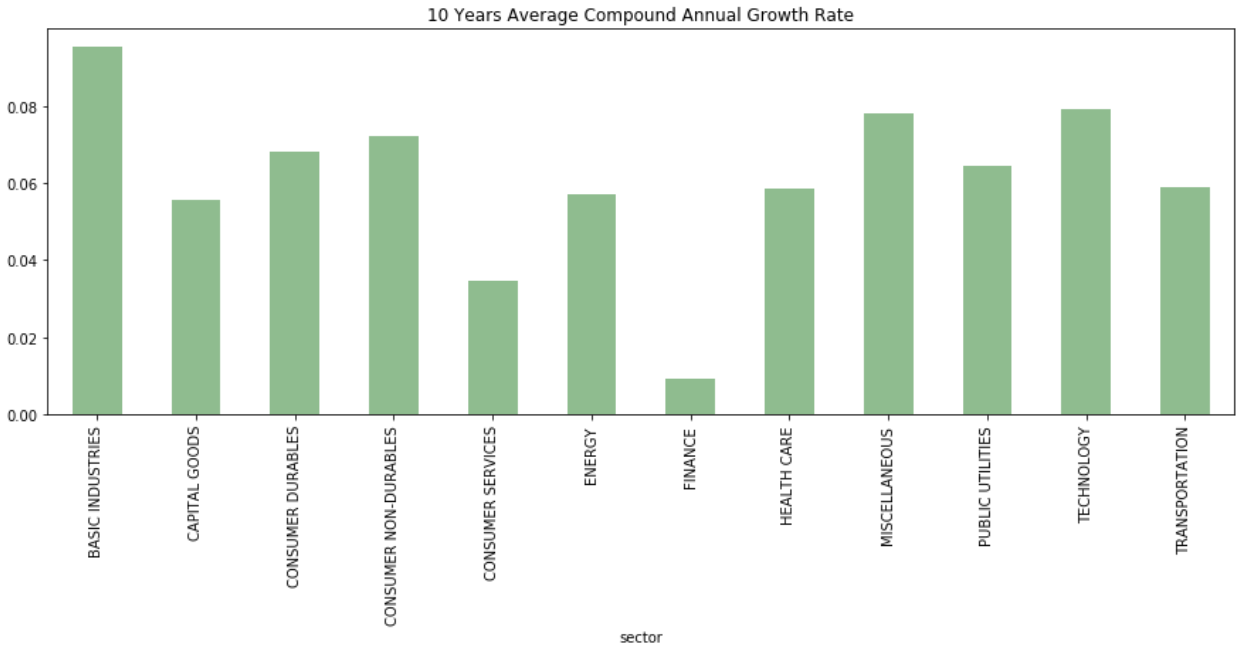
Since CAGR is representing the return for a time period. The one year return makes no sense for CAGR.



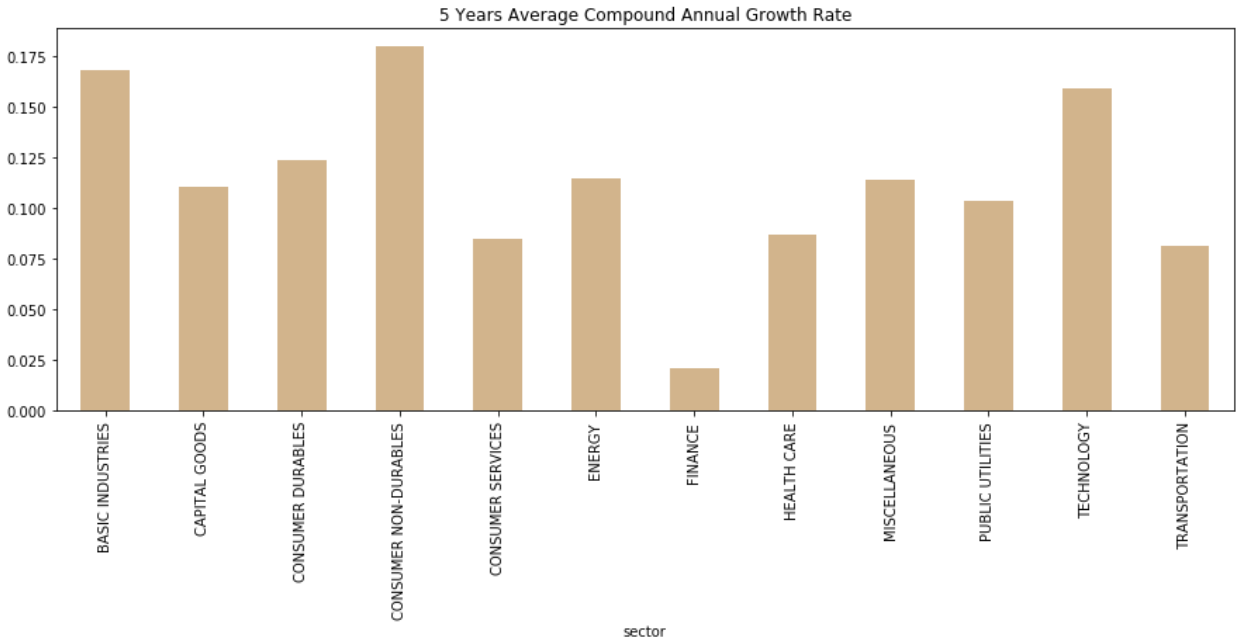
Fifteen Years CAGR:



Ten Years CAGR:



Five Years CAGR:



Limitation of CAGR:

Investment returns are volatile, meaning they can vary significantly from one year to another. However, CAGR does not reflect volatility. CAGR is a pro forma number that provides a "smoothed" annual yield, so it can give the illusion that there is a steady growth rate even when the value of the underlying investment can vary significantly. This volatility, or investment risk, is important to consider when making investment decisions.

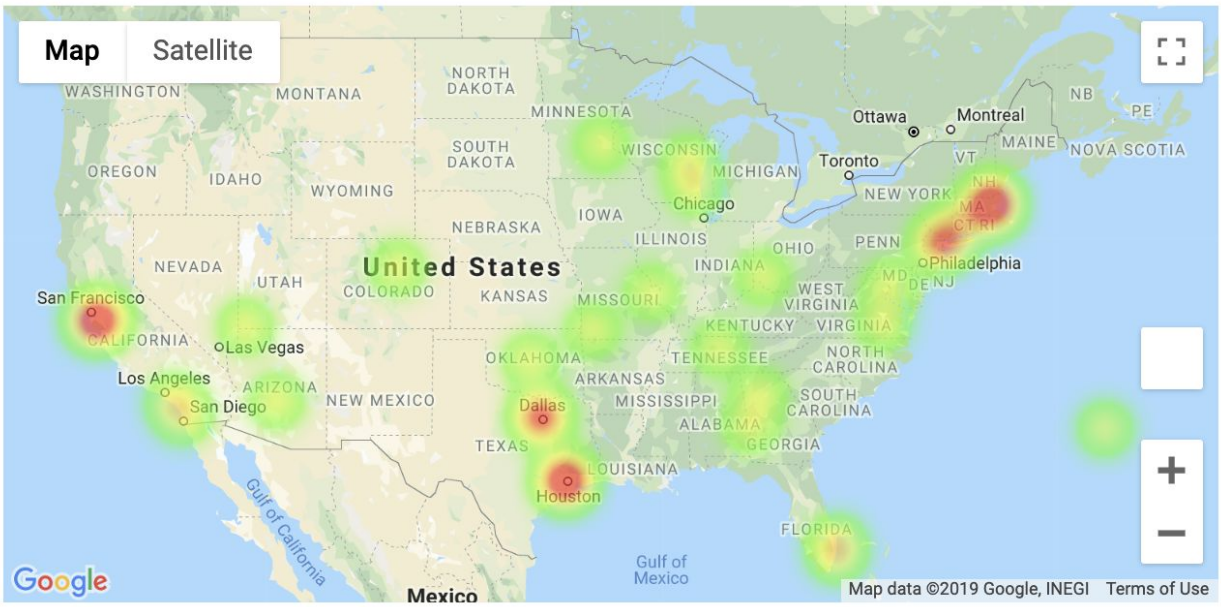
What is a good CAGR:
There is no definition for a good CAGR. However, anything between18%-25% over 5 years is normally considered as a good CAGR. Anything below 12%, it's probably best to choose an alternative investment.

Comparison Of Two Companies:

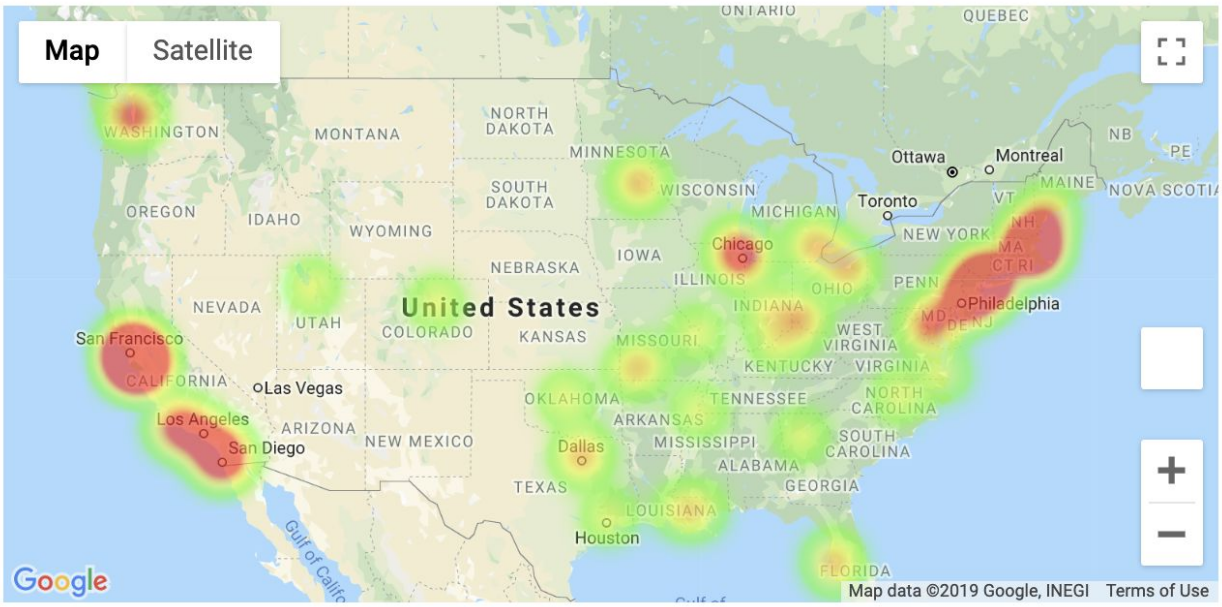
- Are specific areas of the United States more likely to house growing companies? Has this changed over time?
  Living in the bay area, individuals are predisposed to a lot of chatter and marketing surrounding Silicon Valley tech stocks. Often times, these companies gain notoriety by forming quickly, and growing in value even more quickly. However, it's easy to develop a bias and assume that most of the high-growth stocks are in the bay area. Using the Yahoo! Finance and Google GeoCode API's, our team was able to find out.

  We first investigated the stock market over the course of the year 2000. By eliminating stocks that didn't see adequate volume, we then focused in on the opening stock price for each company at the beginning of the year 2000, and compared that to the companies closing stock price at the end of the year 2000 to identify the top 100 stocks who's per-share value grew the most over the year. We then accessed the locations of those stocks and plotted them with a heatmap overlay, shown below.

  In the graph we can see the California hot spots of the Bay Area and Los Angeles being home to some of the hot growth companies of 2000. Aside from Houston and Dallas in Texas, the rest of the growth stocks from 2000 are focused in the Northeast corridor consisting of cities like Philadelphia, Boston, and New York.



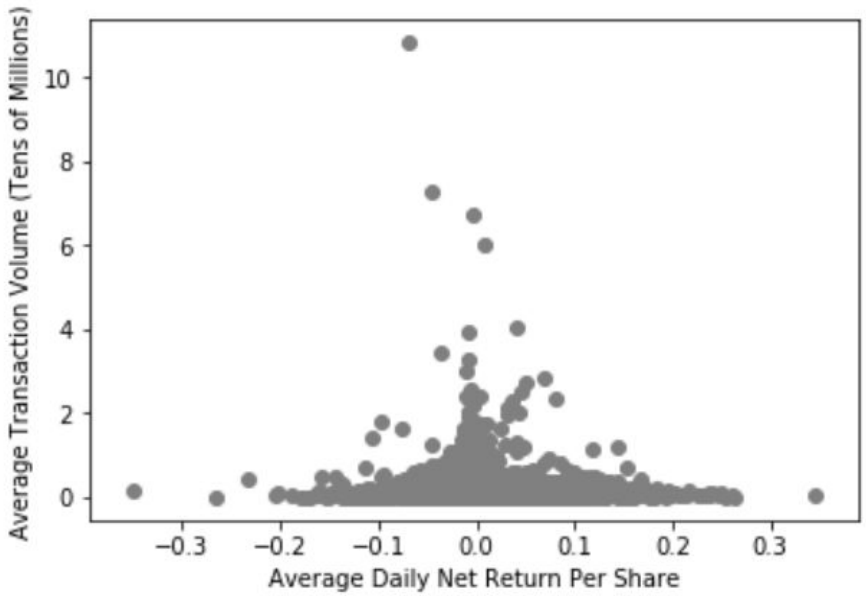
We then compared this with the top 100 growth stocks of 2018 to see how the geographical landscape of growing companies has changed.



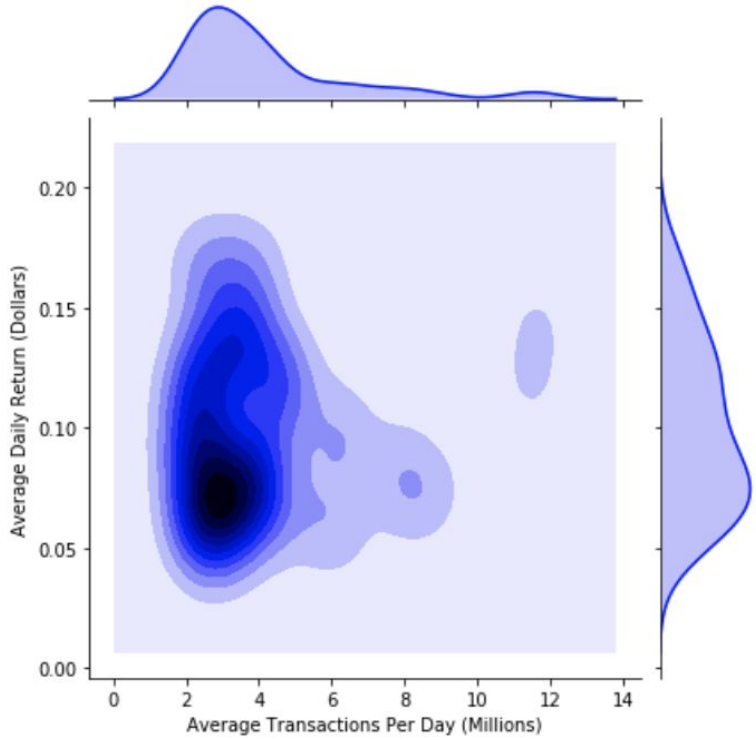
In the year 2018, we can see that the growth stocks continue to be heavily located in the northeast as well as the Bay Area and LA. However, in 2018, we can now see that Seattle in the Pacific Northwest is gaining in popularity, Texas & southern Florida (Miami) have cooled off, while Chicago is now home to some high growth stocks as well.

- Is there a correlation between stock volume and returns?
  Not a day goes by without someone claiming to have a way to beat the market. People are constantly looking for an advantage. Without much financial or stock market

background, it's tempting to just invest money into the stocks and companies you hear about the most. We analyzed the data by calculating for each stock, it's daily gain/loss. That gain/loss per stock was then averaged over the course of the year, as was that stock's volume. The result is a scatter plot as shown below:

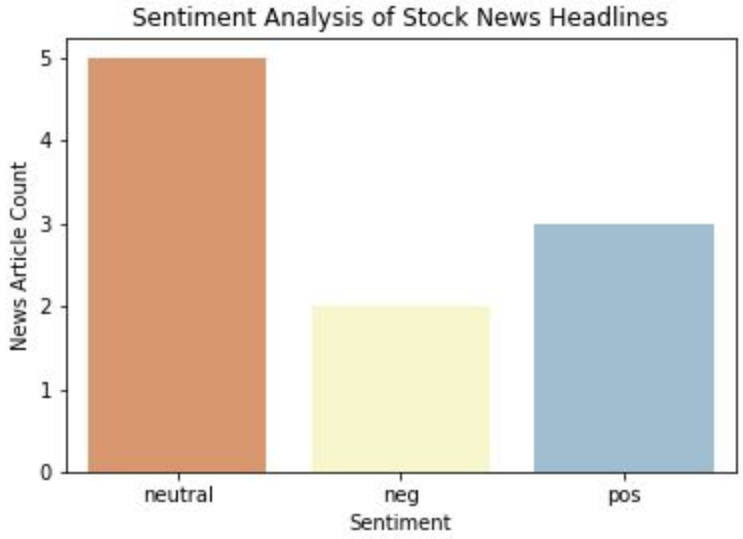


As we can see, there's not much of a correlation that can be disseminated from the data. Many stocks see a net change between -$0.30 and $0.30 per day, with the typical, average transaction volume between 1-4 million. With no correlation found, the question then shifts to: "Which stock should be targeted, based on their volume, for consistent gains?" To answer this, we wanted to look at only the bulk of the data that falls within the normal distribution of share change and volume. After removing outliers, we then plotted the data using a Kernel Density Estimation graph.



We can see, after removing the outliers, that if an individual were to focus on popular (but not the most popular) stocks, which see daily volume between 2-4 million, that the user should expect to see an average daily return of $0.05 - $0.10 cents per share.

- Can digital media be used to gauge sentiment of a particular stock in order to guide the stock owner in buying more, selling, or holding current shares?
  In the age of digital media, access to information is as easy as it has ever been. The data community has now published several scientific papers on ways that media and news coverage (and potentially fake news) can influence a stock's day-to-day value. This lends to the idea that if the news headlines circulating about a given stock can be judged based on their sentiment, it may clue in the shareholders to being proactive in what they should decide to do with their investments. To do this, we built code to prompt a given user to enter a stock ticker. We then take that stock ticker, and query the Yahoo! Finance API to return up to the 10 most recent news headlines related to that stock. We then query a sentiment analysis API to return the label of 'positive', 'neutral', or 'negative' for a given stock's headlines, and graph the resulting grades on a bar chart as shown below, for the stock ticker 'AMZN' (Amazon).


Sentiment Analysis of Stock News Headlines
News Article Count
0
1
2
3
4
5
neutral
neg
pos
Sentiment