**Exploratory Data Analysis**

**Prediction of sector performance during recessions**

**By: Ash Tao, Chris Burley,**

**Maica Tran and Sabrina Tsan**

Table of Contents

[1. INTRODUCTION 3](#_Toc99289637)

[Aim 3](#_Toc99289638)

[2. STEP BY STEP APPROACH 3](#_Toc99289639)

[Data Fields and Variables – 3](#_Toc99289640)

[Approach to breakdown the problem 4](#_Toc99289641)

[**What we already know (Baseline knowledge, research, and studies):** 4](#_Toc99289642)

[**Cleansing Data** 4](#_Toc99289643)

[**Analysis** 4](#_Toc99289644)

[**Time-based analysis** 4](#_Toc99289645)

[Results - Analysis 5](#_Toc99289646)

[**1. We can tell some trends from the cause of a recession** 5](#_Toc99289647)

[**2. Trade volume makes the performance differences:** 6](#_Toc99289648)

[**3. the previous price can help to make a judgment of buying or selling:** 11](#_Toc99289649)

[**4. The correlation between sectors become stronger than ever:** 13](#_Toc99289650)

[**5. Continue holding or buying energy sector doesn't necessarily mean dangerous:** 15](#_Toc99289651)

# INTRODUCTION

## **Aim**

Try to predict the sector trend in a recession for the S&P 500 market. The S&P 500 includes 500 major U.S. public companies focused primarily on market capitalization.

The S&P 500 is widely regarded as one of the best indicators of large-cap U.S. stocks and the broader stock market.

Use Panda to get S&P50 sector data for the last four Yahoo Finance recessions and visualize all aspects of it. Finally, let's look at some ways to analyse stock risk based on previous performance history.

# STEP BY STEP APPROACH

## **Data Fields and Variables –**

**yfinance**: a python package that downloads market data from Yahoo! Finance's API

**yf.download()**: download the data based on the selected period.

*Each period will have a separately DataFrame.*

|  |  |
| --- | --- |
| Fields | Fields Information |
| Date | The date for this record |
| Open | The opening price on that date |
| High | The highest price on that date |
| Low | The lowest price on that data |
| Close | The closing date on that date |
| Adj Close | Amends a stock's closing price after accounting for any corporate actions. |
| Volume | Traded volume on that date |

**sp500\_companies.csv:** Details of 500 S&P companies. (Date - 20220316)

|  |  |
| --- | --- |
| Fields | Field Information |
| Symbol | Id for the stock |
| Shortname | Name of the stock |
| Longname | Name of the stock |
| Sector | Sector the stock under |
| Industry | Industry the stock under |
| Currentprice | Current Price: The latest price for the stock |
| Marketcap | The total value of all a company's shares of stock. |
| Ebitda | A stand for earnings before interest, taxes, depreciation, and amortization |
| Revenuegrowth | The amount of money the company makes over a pre-determined time compared to the previous, identical amount of time. |
| City | The register city of the stock |
| State | The state of the stock |
| Country | The city of the stock |
| Fulltimeemployees | Full Time Employees |
| Longbusinesssummary | Business Summary of the company of the stock |
| Weight | The percentage value of each stock in the portfolio |

**Feature Table:** Feature Table Trying to calculate the weight of each sector

|  |
| --- |
| Fields |
| Sector |
| Marketcap |

## **Approach to breakdown the problem**

### **What we already know (Baseline knowledge, research, and studies):**

* During the recession, the total market index will decrease.
* “Daily Return” is a concept for analysing index performance, a daily change in price as a percentage of the opening price. *Formula: (Closing price - opening price)/opening price*
* Moving Average: a simple [technical analysis](https://www.investopedia.com/terms/t/technicalanalysis.asp) tool that smooths out price data by creating a constantly updated [average price](https://www.investopedia.com/terms/a/averageprice.asp). The average is taken over a specific period, like days, minutes, weeks, or any period the trader chooses.
* Expected Return & Rick can be displayed on a scatter plot based on the mean of daily return and standard deviation of daily return.

### **Cleansing Data**

* The data retrieved from Yahoo finance don’t need to be cleaned.
* There is no missing data in the column “Sector” & “Marketcap”, CSV files don’t need to be cleaned as well.

### **Analysis**

* 4 recessions period:
* Covid-19 from 2020,2,3 to 2020,4,1
* Financial crisis (GFC) from 2007,12,3 to 2009,6,1
* Dot-Com bubble from 2001,3,1 to 2001,11,1
* Find out the correlation between the data – trying to understand what correlation of each dimension, which is including closing price, price return, daily trade volume and moving average.
* See how dimensions affect each other and see if dimension could be affecting the sector performance.
* We'll be answering the following questions along the way:

1. Can we tell the trend from the cause of the recession?
2. What makes one sector perform differently from the others?
3. Can the previous price help to make a judgment of buying or selling?
4. Is there a correlation between sectors?

### **Time-based analysis**

* Find out sector price over days, sector return over days, daily return over days, daily trade volume over days and moving average over days.

## **Results - Analysis**

### **1. We can tell some trends from the cause of a recession**

**1.1 The performance of each sector varies, some sectors gain, and some lose.**

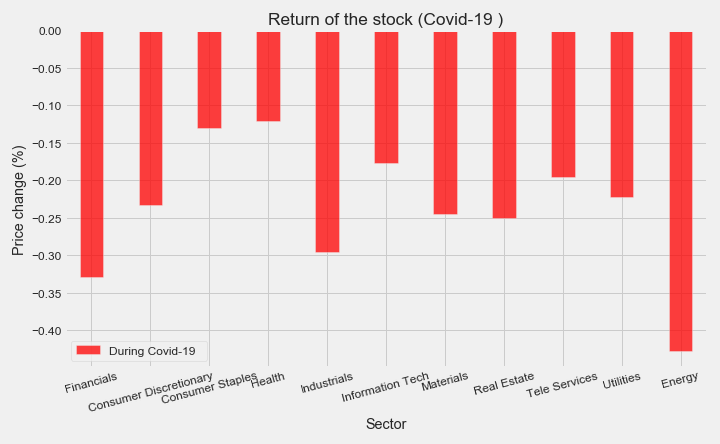
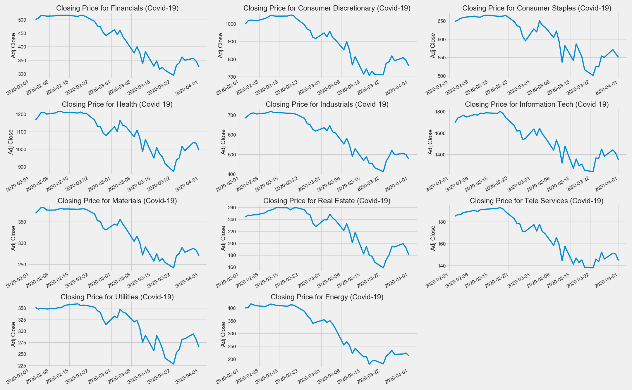
International events can lead to market recessions. Because of the different nature of events, the market will behave differently during a recession.

Some sectors related to the causes of the recession may have the biggest losses and become disaster areas. (e.g., the financial sector experienced the largest losses in 2008 GFC).

And some may be affected less during the recession. (e.g., the health sector in covid, has the least loss).

Furthermore, people's predictions about the future will also affect the current market.

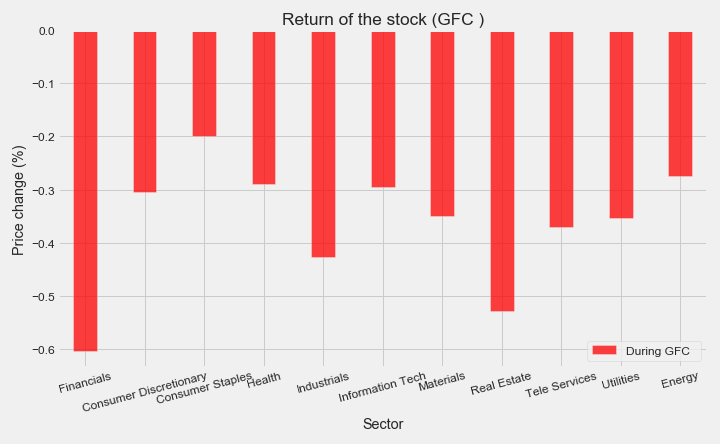
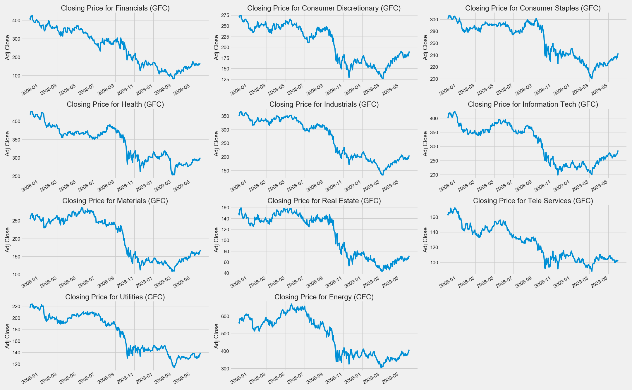
* **COVID-19 – Feb 20 to Apr 20**



All sectors fell, the energy sector fell the most, and the health sector fell the least.

The recession is caused by a global pandemic.

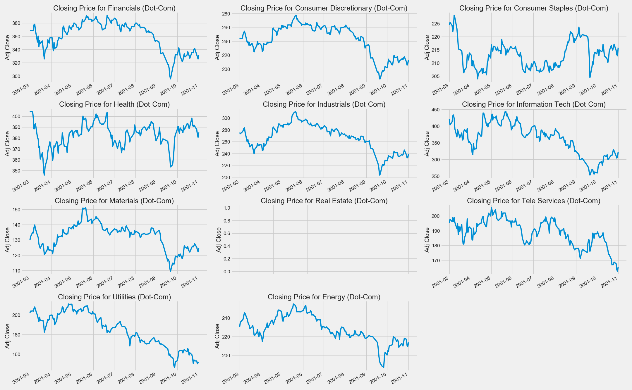
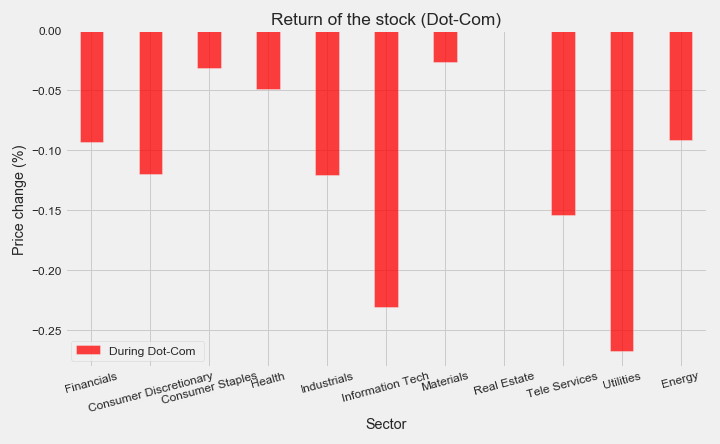
* **The financial crisis (GFC) – Dec 07 to Jun 09**



All sectors fell, the financial sector fell the most, and the health fell the least.

The financial crisis was primarily caused by deregulation in the financial industry.

* **Dot-Com bubble – Mar 01 to Nov 01**

All sectors fell, Utilities fell the most, and Materials fell the least. No information on Real Estate backs that time.

We think the most fell of Tech and the least fell is Materials makes sense. But we are not sure what happened on the Utilities. Why the price dropped greater than the tech sector?

**1.2 The causes of a recession can help predict the trend of the market.**

Based on the causes of the recession, people can predict a little bit sector’s performance regarding the cause of the recession.

**1.3 Limitation**

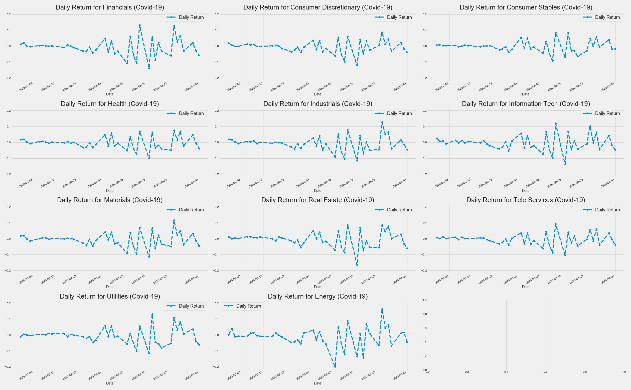
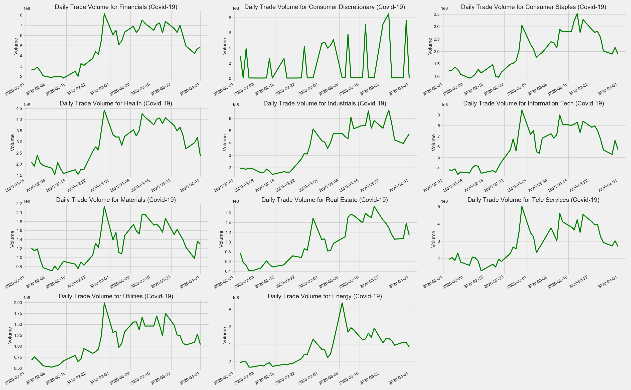
However, the cause is often discovered after the market has crashed due to the long tail effect of the market. Therefore, the judgment only based on the cause will not make a big difference.

### **2. Trade volume makes the differences on sector performance:**

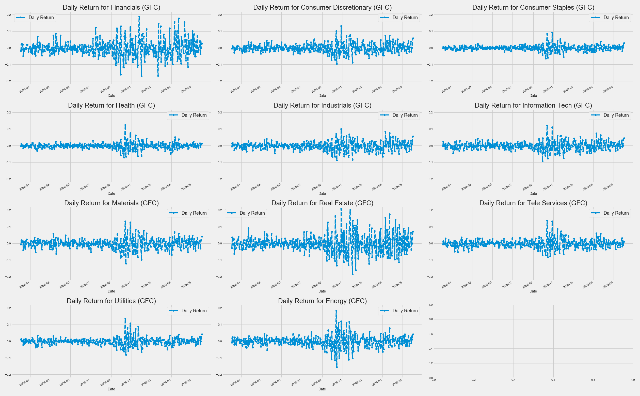
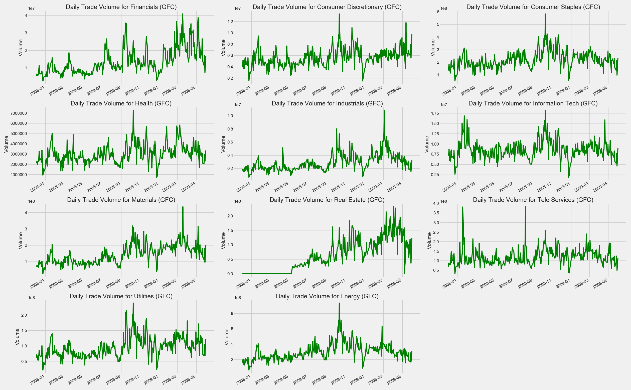
**2.1 Trade volume reflects market confidence and will affect the price.**

Volume measures the number of shares traded in a sector. Volume can indicate market strength. From the plot we generated, we found that “Daily Trade Volume" and "Daily Retune" seem to have similar patterns through observation.

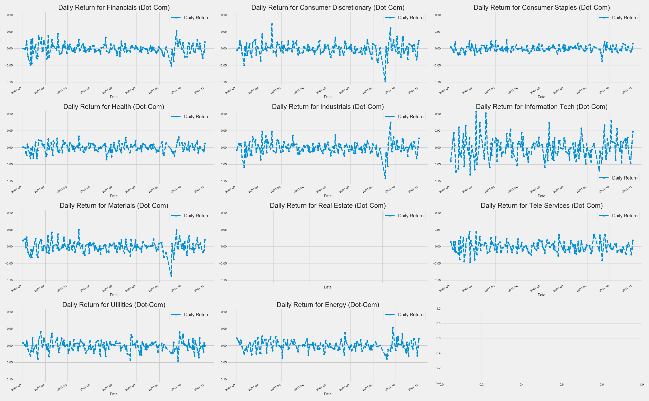
* **COVID-19 – Feb 20 to Apr 20**

* **The financial crisis (GFC) – Dec 07 to Jun 09**

* **Dot-Com bubble – Mar 01 to Nov 01**



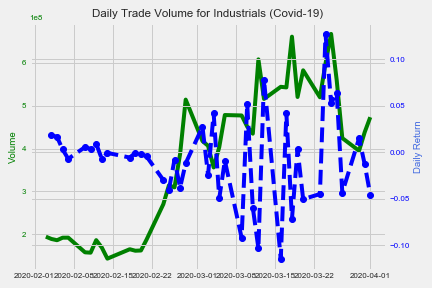
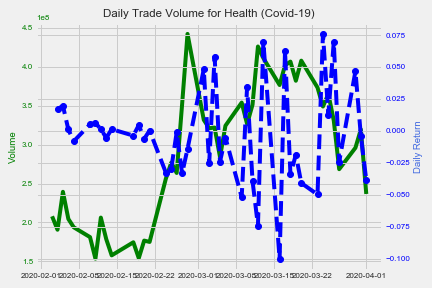
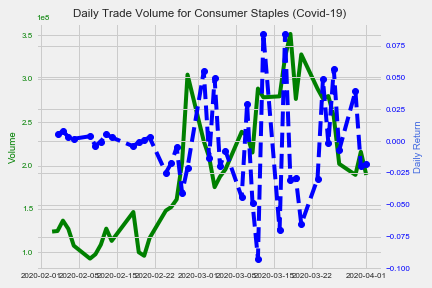
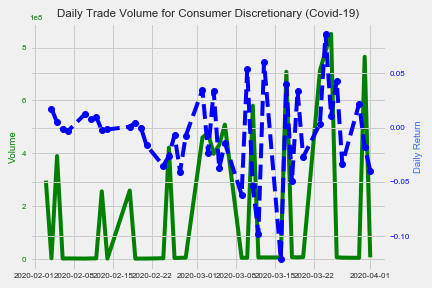
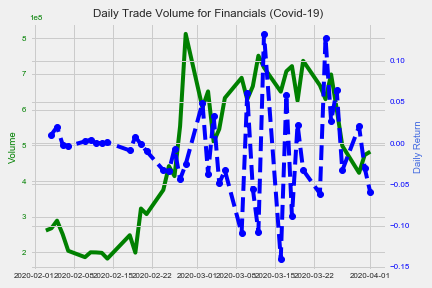
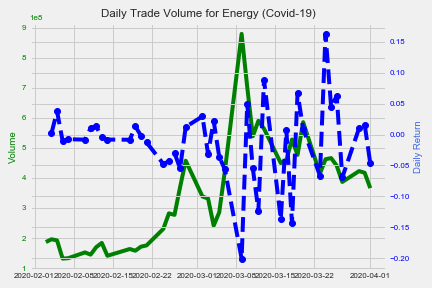
There is no daily trade volume data during the Dot-Com bubble.

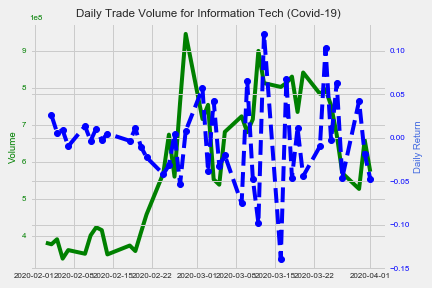
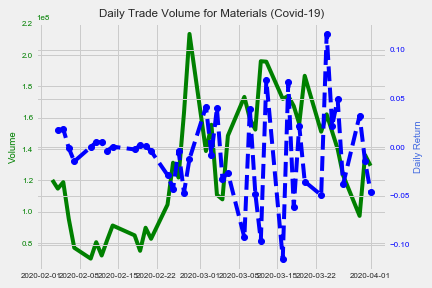
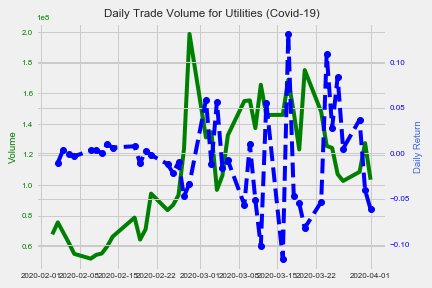
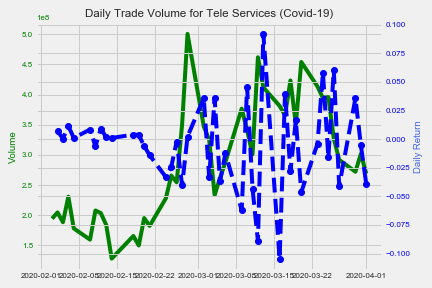
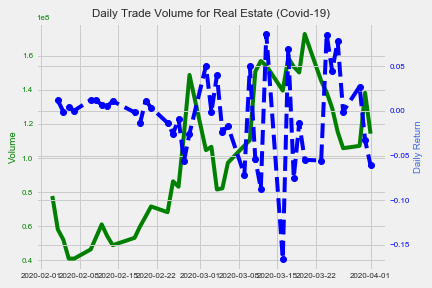
**2.2 There is a pattern between Daily Return and Daily Volume**

The peaks and deeps on Return & Volume always seem to appear in pairs.

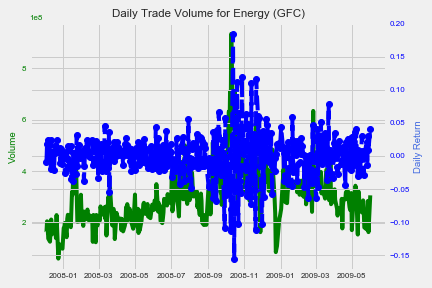
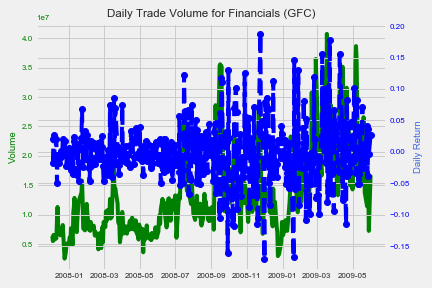
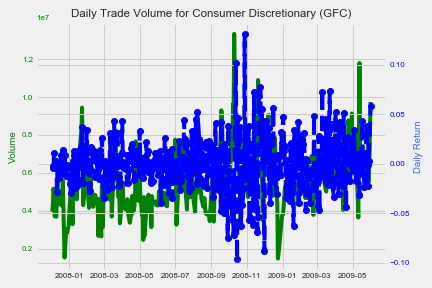
The turning point of volume (the direction change of acceleration) often indicates the turning point of return.

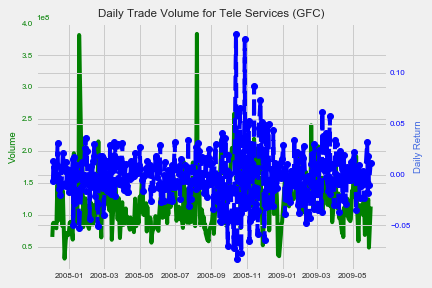
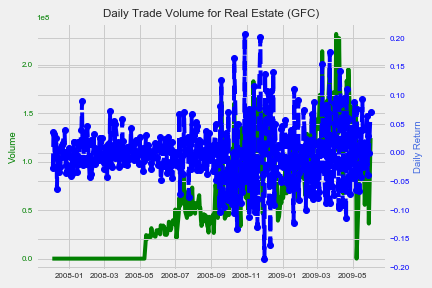
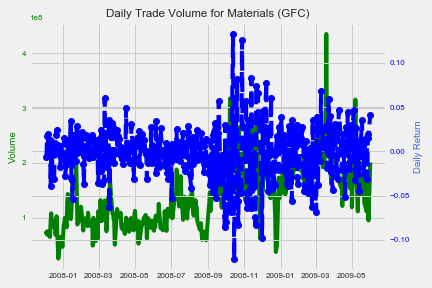
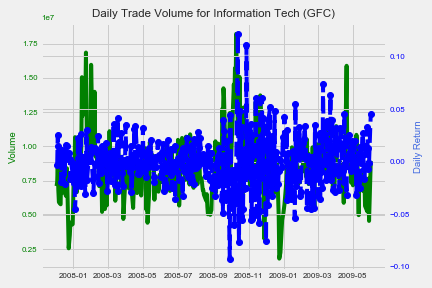
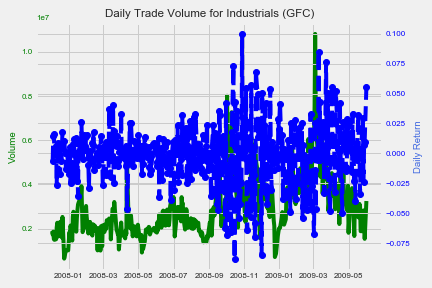
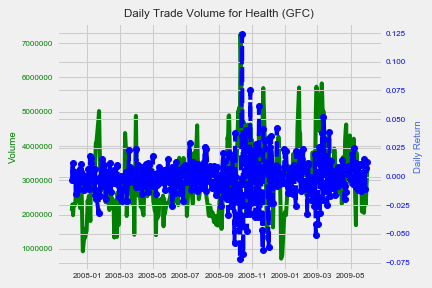
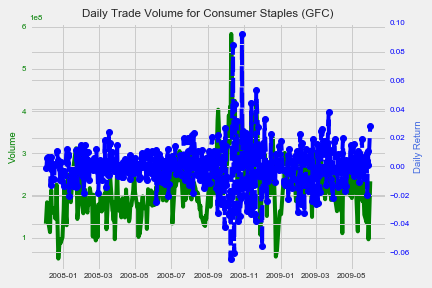
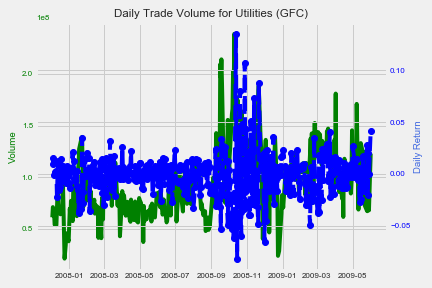
* **COVID-19 – Feb 20 to Apr 20**



****

* **The financial crisis (GFC) – Dec 07 to Jun 09**

**  **

** **

* **Dot-Com bubble – Mar 01 to Nov 01**

There is no daily trade volume data during the Dot-Com bubble.

**2.3 Daily Return and Daily Volume can help predict the trend of the market.**

Daily Return and Daily Volume are closely related due to their definitions. The daily trading volume affects the price, and the daily return is obtained by the price. Mathematically speaking, there is a correlation between these two.

From a market perspective, trading volume can help investors identify market trends.

If the volume keeps moving on one way, the price usually does not change its direction. In other words, if the volume trend remains unchanged (continuously increasing or decreasing), the price trend does not change. At this point, it is safe to keep the strategy you have.

If the volume trend changes, from increasing to decreasing (or a decrease becomes an increase), which means that the price is facing a reversal point. At this point, you may need to consider changing your strategy.

**2.4 Limitation**

However, in theory, the above statement is correct. But in the real world, this theory might not work due to various reasons. For example, volume changes every second, and we can only know the change of trend after it crosses the extreme value. Therefore, the relationship between Volume and Return is not enough to help investors make a clear judgment. But there is indeed a close correlation between them, and volume can reflect investors' confidence and interest in the market.

### **3. the previous price can help to make a judgment of buying or selling:**

**3.1 The moving average (MA) can help smooths out price data and help the judgment.**

As can be seen from the plot below, the daily closing price moves up and down around the average line and make crossovers. The different average lines also crossed.

The moving averages can help to identify trend direction and to determine support and resistance levels.

The short-term average is trending closer to price.

* **COVID-19 – Feb 20 to Apr 20**

Diagram

Description automatically generated

* **The financial crisis (GFC) – Dec 07 to Jun 09**

Chart, map

Description automatically generated

* **Dot-Com bubble – Mar 01 to Nov 01**

*No real estate price data during the Dot-Com period. The plot of this sector is blank*

Chart

Description automatically generated

**3.2 The moving average (MA) can help predict the trend**

Crossovers are one of the main moving average strategies.

First, we can judge by the crossover on the closing price and the moving average line. When the price crosses above or below a moving average, there will be a signal for change in trend.

Chart, map

Description automatically generated

Second, we can judge by the crossover on the short-term and long-term moving average lines. When the shorter-term MA crosses above the longer-term MA, it's a buy signal, as it indicates that the trend is shifting up. This is known as a golden cross. Meanwhile, when the shorter-term MA crosses below the longer-term MA, it's a sell signal, as it indicates that the trend is shifting down.

Chart, line chart

Description automatically generated

**3.3 Limitation**

Moving averages are calculated based on historical data and nothing about the calculation is predictive in the real world. if the price becomes choppy, the price may swing back and forth, generating multiple trend reversals or trade signals. When this occurs, we need another indicator to help clarify the trend.

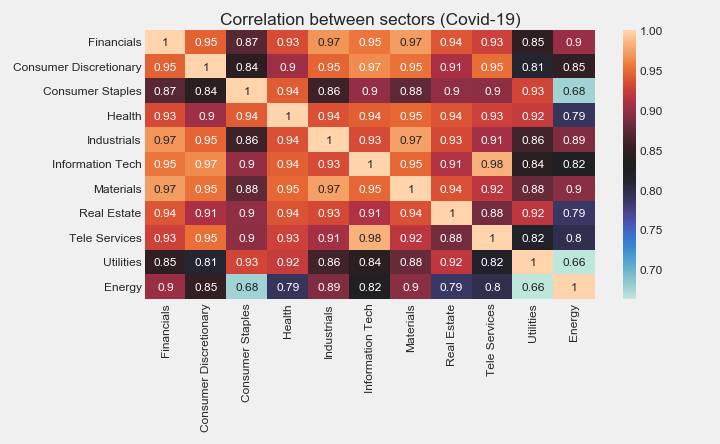
If prices start fluctuating, sometimes the market won't respect MA support/resistance trade signals. Moving averages work well in strong trending conditions but poorly in ranging conditions.

### **4. The correlation between sectors become stronger than ever:**

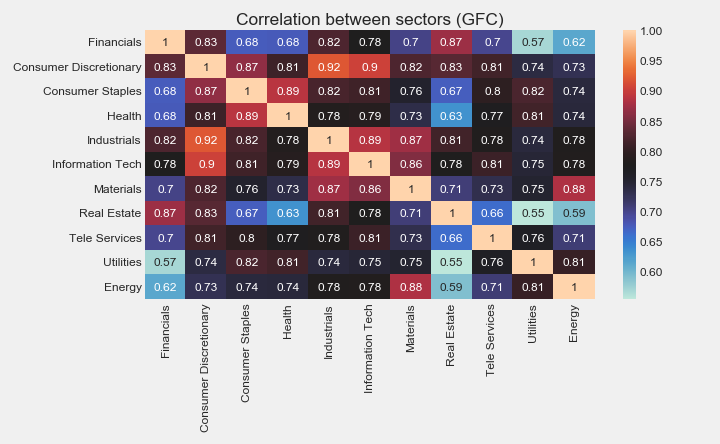
The correlations between the various industries are getting tighter over time. The colour of the plot is getting warmer.

The colour of the heat map is getting warmer means that all industries are getting closer than ever due to the Cross-Industry. Thus, during the recession, no matter which sector you invest in, it will be affected strongly by others.

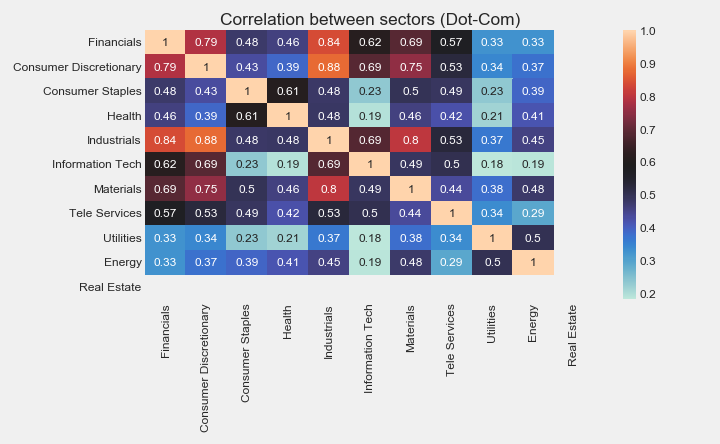
* **COVID-19 – Feb 20 to Apr 20**



* **The financial crisis (GFC) – Dec 07 to Jun 09**



* **Dot-Com bubble – Mar 01 to Nov 01**



To understand it, let's use Apple as an example. If Apple sells strongly on its newly released products (iPhone, Mac Book, etc.), then all companies in its supply chain, such as mobile phone accessories manufacturers or retailers, will have a good business performance as well. Their stock price will go up, and their sector is different from Apple.

Therefore, when making investment decisions, the correlation between sectors should be a very important consideration. Regarding this topic, other data need to be introduced from different perspectives. We won’t be able to discuss it more deeply here.