FE-582 Foundations of Financial Data Science, Spring 2021

Group 5: Daniel Bachalis, Gurjivan Kalkat, Naveen Nagarajan

Forecasting stock prices during pandemic

Introduction:

The current pandemic has affected global markets and economies tremendously, especially in the United States. The research being performed will look at the performance and correlations across commodities, macroeconomic indicators, equity indices, COVID 19 data, and TSA data before and after the pandemic. Certain assets and commodities performed better post pandemic while others are still lagging in terms of returns. An initial model will be created starting from the beginning of 2019 to the start of the pandemic. The model will then be changed to account for the current environment. Based upon the analysis of this data, a determination will be made as to the best asset allocation for investors and what can be expected going forward.

Research statement (questions):

We will use data science techniques to do the following:

- Predict which business cycle we are currently in and how long it will last as it relates to the Covid-19 pandemic.
    - Develop Covid-19 prediction model to leverage for pandemic duration prediction.
    - Compare Data beginning from the pre pandemic levels starting from 2019 until March 2021, to observe how different asset classes and macroeconomic data has been affected by Covid-19
- Make recommendations for portfolios and asset allocation in order to maximize returns and hedge against risk.

Data:

```
> head(COVID_BTC_SPX_DOW_MAC)
# A tibble: 6 x 14
  Date           TSA    Cases  Rate Deaths Bitcoin   TLT   DJIA    NDX  Gold   Oil   SPX   GDP Unemployment
  <date>       <dbl>    <dbl> <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>        <dbl>
1 2021-03-01 1047418 28461597  140.  514521  49249.  141. 31536. 13589. 1725.  60.6 3902.    NA           NA
2 2021-02-26 1096348 28290518  142.  510458  45752.  143. 30932. 13192. 1734.  61.5 3811.    NA           NA
3 2021-02-25 1051149 28217301  142.  508114  48291.  139. 31402. 13119. 1771.  63.5 3829.    NA           NA
4 2021-02-24  802230 28143368  140.  505803  48745.  141. 31962. 13598. 1805.  63.2 3925.    NA           NA
5 2021-02-23  714725 28068247  138.  502594  48173.  142. 31537. 13465. 1806.  61.7 3881.    NA           NA
6 2021-02-22  963280 27998447  134.  500236  54182.  142. 31522. 13533. 1810.  61.7 3876.    NA           NA
```

```
> tail(COVID_BTC_SPX_DOW_MAC)
# A tibble: 6 x 14
  Date           TSA Cases  Rate Deaths Bitcoin   TLT   DJIA   NDX  Gold   Oil   SPX   GDP Unemployment
  <date>       <dbl> <dbl> <dbl>  <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>        <dbl>
1 2019-01-09 1739642    NA    NA     NA    3995.  121. 23879. 6957. 1294.  55.3 2585.    NA           NA
2 2019-01-08 1733739    NA    NA     NA    3993.  121. 23787. 6897. 1285.  53.5 2574.    NA           NA
3 2019-01-07 2044043    NA    NA     NA    3990.  122. 23531. 6823. 1289.  52.4 2550.    NA           NA
4 2019-01-04 2150571    NA    NA     NA    3812.  122. 23433. 6739. 1286.  52.2 2532.    NA           NA
5 2019-01-03 2202111    NA    NA     NA    3796.  124. 22686. 6464. 1294.  51.7 2448.    NA           NA
6 2019-01-02 2345103    NA    NA     NA    3882.  122. 23346. 6666. 1285.  51.2 2510.    NA           NA
>
```

```
# A tibble: 6 x 14
  Date           TSA    Cases  Rate Deaths Bitcoin   TLT   DJIA    NDX  Gold   Oil   SPX    GDP Unemployment
  <date>       <dbl>    <dbl> <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl>  <dbl>        <dbl>
1 2020-12-31  874406 19958545 394.  345737  29112.  158. 30606. 12888. 1898.  48.7 3756. 21488.         6.77
2 2020-09-30  634046  7237343  88.7 206918  10742.  163. 27782. 11168. 1886.  41.8 3363  21170.         8.8
3 2020-06-30  500054  2627564  87.2 127410   9148.  164. 25813. 10059. 1781.  40.1 3100. 19520.        13.1
4 2020-03-31  146348   185812  39.5   5367   6446.  165. 21917.  7700. 1577.  35.5 2585. 21561.         3.8
5 2019-12-31 2392331       NA  NA       NA  7180.  135. 28538.  8973. 1517.  55   3231. 21747.         3.6
6 2019-09-30 2082179       NA  NA       NA  8241.  143. 26917.  7999. 1472.  50.1 2977. 21540.         3.6
> 
```

Datasets were retrieved from bloomberg [3]

Exploratory Data Analysis (EDA):

The following figures below display the initial summary statistics for all of the data:

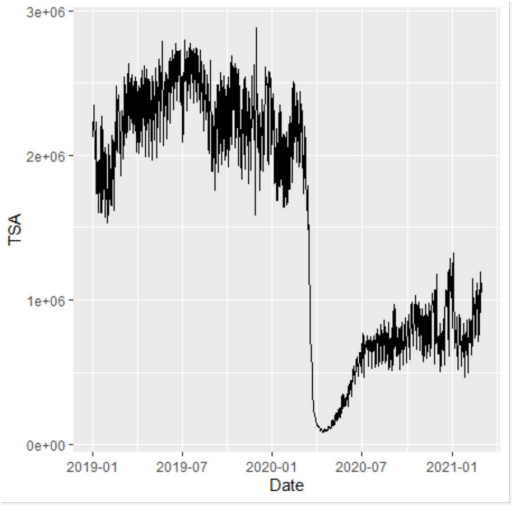
```
                            TSA                Cases           Rate              Deaths
nbr.val              790.0000000            404.000000   404.0000000          404.0000000
nbr.null               0.0000000              0.000000    38.0000000           37.0000000
nbr.na                 0.0000000            386.000000   386.0000000          386.0000000
min                87534.0000000              1.000000     0.0000000            0.0000000
max              2882915.0000000       28461597.000000   525.8000000       514521.0000000
range            2795381.0000000       28461596.000000   525.8000000       514521.0000000
sum           1215601605.0000000     3243068853.000000 59624.9300000     72673304.0000000
median           1844344.0000000        5161216.000000    91.9650000       165278.5000000
mean             1538736.2088608        8027398.150990   147.5864604       179884.4158416
SE.mean            30620.0461417         434257.586392     7.1665658         7088.5380846
CI.mean.0.95       60106.3913407         853693.063190    14.0885218        13935.1297033
var          740693908317.4505615 76186179140904.109375 20749.3046641 20299938359.5636177
std.dev          860635.7582145        8728469.461532   144.0461893       142477.8521721
coef.var              0.5593134              1.087335     0.9760122            0.7920522
                        Bitcoin
nbr.val             790.0000000
nbr.null              0.0000000
nbr.na                0.0000000
min                3360.5340759
max               57128.6426065
range             53768.1085306
sum             9170259.9507282
median             9279.2832011
mean              11607.9239883
SE.mean             332.1773335
CI.mean.0.95        652.0558692
var           87170006.9169658
std.dev            9336.4879327
coef.var              0.8043202
```

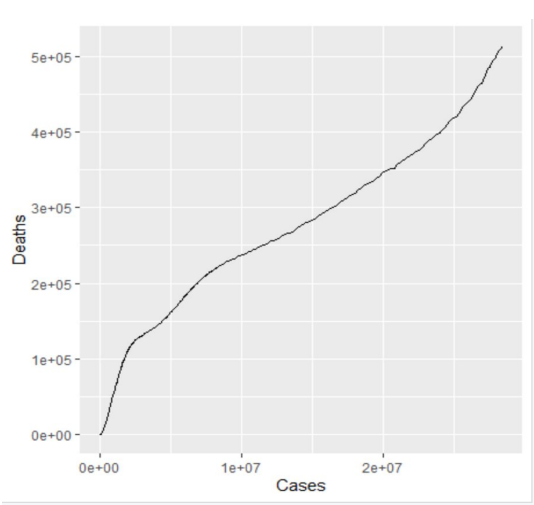
	TLT	DJIA	NDX	Gold	Oil
nbr.val	544.0000000	544.00000000	544.0000000	544.0000000	544.0000000
nbr.null	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000
nbr.na	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000
min	118.6600000	18591.93000000	6463.5040000	1270.6900000	29.6300000
max	171.5700000	31961.86000000	14095.4730000	2063.5400000	63.5300000
range	52.9100000	13369.93000000	7631.9690000	792.8500000	33.9000000
sum	79432.1800000	14662531.85000000	5107001.3730000	871572.9600000	26476.9600000
median	144.8050000	26806.95500000	8641.2270000	1560.5600000	51.3800000
mean	146.0150368	26953.18354779	9387.8701710	1602.1561765	48.6708824
SE.mean	0.6763470	94.63855349	81.2689852	9.6788159	0.3210640
CI.mean.0.95	1.3285772	185.90252300	159.6401131	19.0125085	0.6306796
var	248.8502547	4872311.95902883	3592928.4865166	50961.6358668	56.0766515
std.dev	15.7749883	2207.33141124	1895.5021726	225.7468402	7.4884345
coef.var	0.1080367	0.08189502	0.2019097	0.1409019	0.1538586

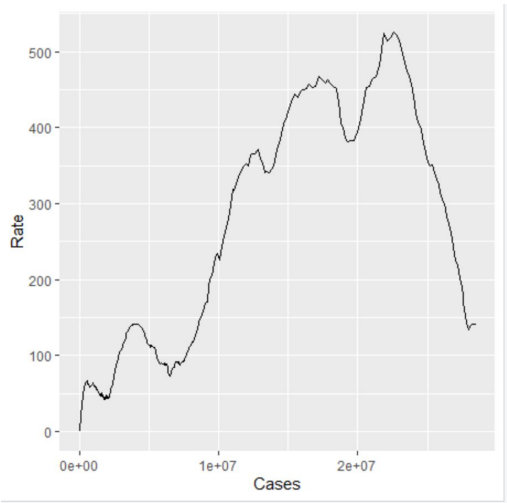
	GDP	Unepployment
nbr.val	9.00000000	9.0000000
nbr.null	0.00000000	0.0000000
nbr.na	0.00000000	0.0000000
min	19520.11000000	3.6000000
max	21747.39000000	13.0700000
range	2227.28000000	9.4700000
sum	190382.15000000	51.0100000
median	21329.88000000	3.8300000
mean	21153.57222222	5.6677778
SE.mean	221.75142129	1.1084220
CI.mean.0.95	511.35969448	2.5560258
var	442563.23559444	11.0573944
std.dev	665.25426387	3.3252661
coef.var	0.03144879	0.5866966

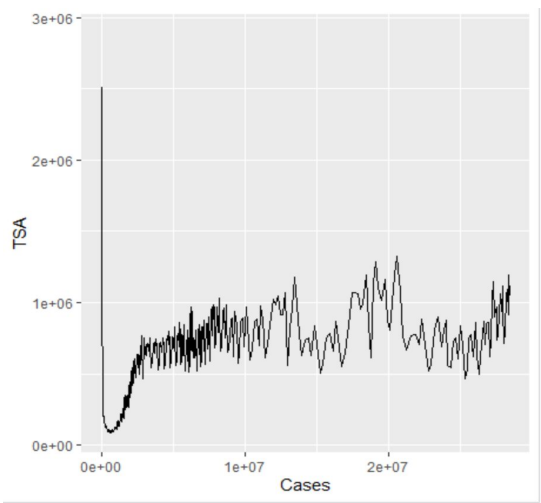
The following graphs and plots attempt to look at multiple data sets and make predictions about how the assets are correlated to each other and how COVID 19 changed these correlations:

TSA and COVID 19 Comparison





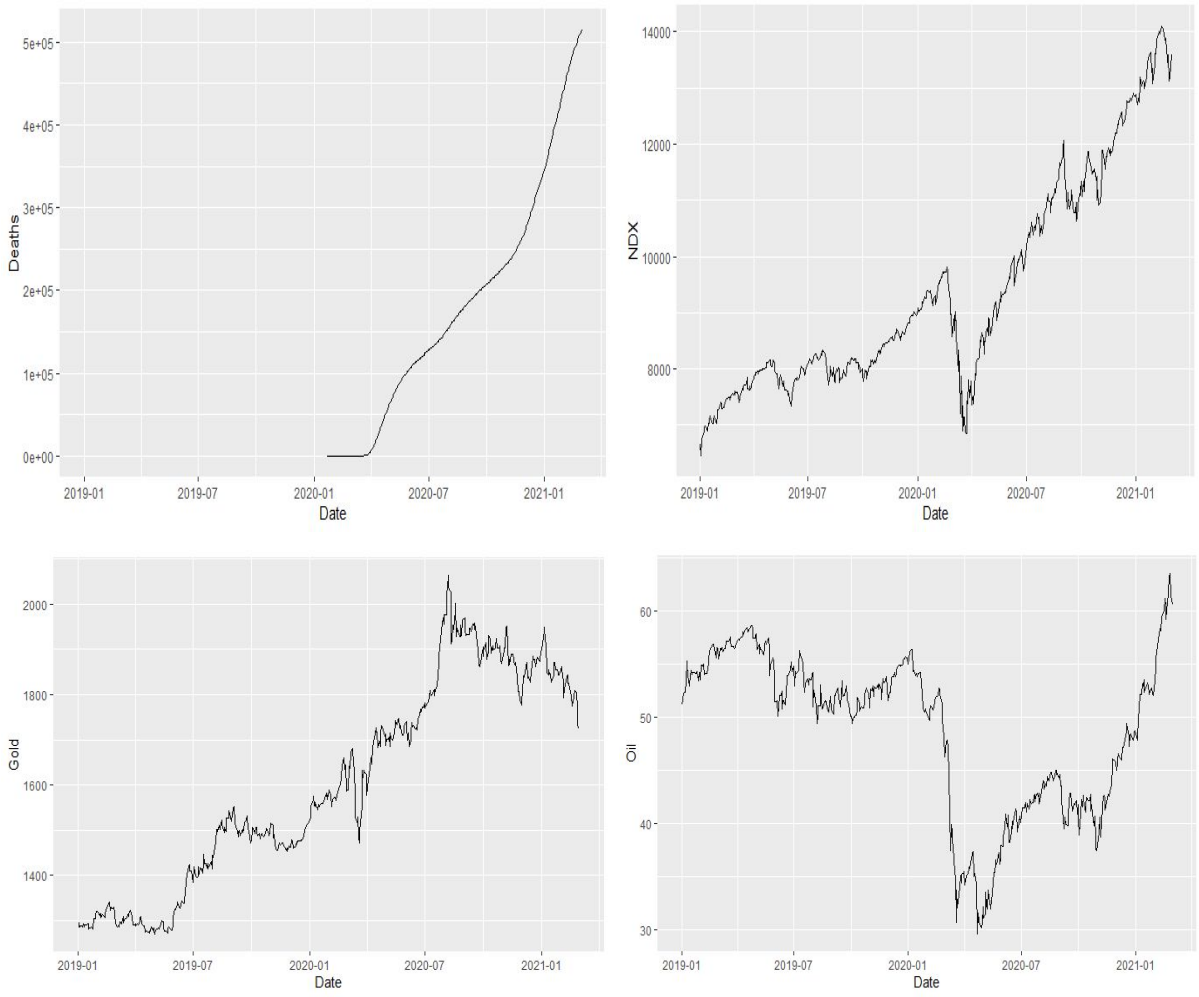




Based on the above initial data, it is clear to see that the amount of people passing through TSA at the airport significantly decreased as a result of COVID 19. The amount of people passing through TSA immediately began to pick up as travel restrictions began to become lifted, however numbers are still well below pre pandemic levels. In addition it is clear to see that there is a positive correlation between the total number of cases and the total number of deaths. However, as the Cases rose above approximately 2.5 million, the death rate per 100,000 cases started to decrease significantly, which could be indicative of several things: treatment of the symptoms improved as healthcare workers gained more knowledge and experience with cases, there were more resources mobilized to treat cases, the virus had potentially mutated to a less deadly strain, the susceptible/exposed population is more resilient (i.e. older/at risk population that was going to be exposed is already dead and only younger/less risky population is still being exposed).

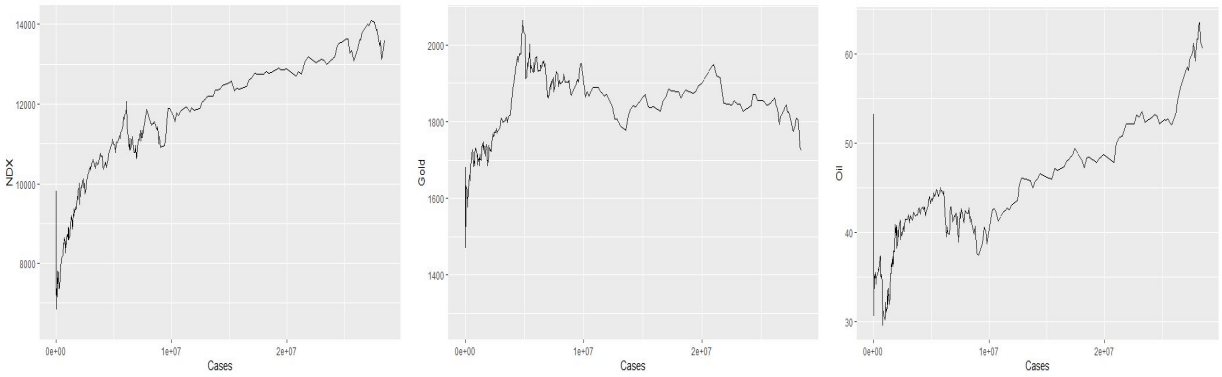
Nasdaq, Gold and Oil comparison with Covid 19 Data

Covid-19 Deaths, NDX, Gold and Oil Prices over Time

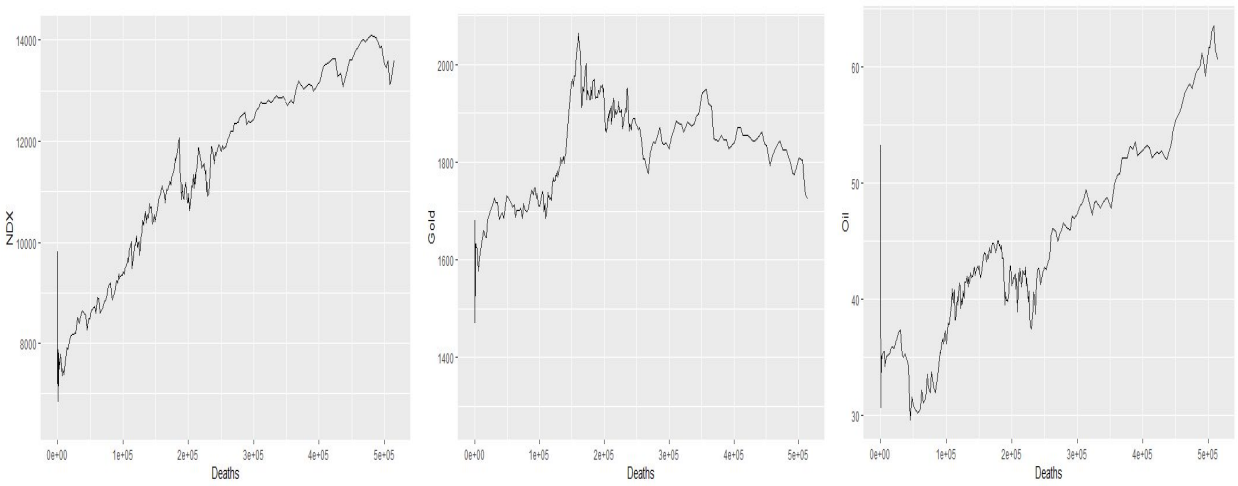


There is a clear dip in the Nasdaq (NDX) and Oil prices at the start of the Covid-19 pandemic. There is also a small, brief dip in Gold price as well, but it rapidly recovers. At this point, the NDX and Oil prices have recovered as well. All three prices seem to have maintained their original trend. This could lead to the hypothesis that all three prices are not positively correlated to Covid-19 deaths, at least not directly.

NDX, Gold and Oil Prices vs. Covid-19 Cases



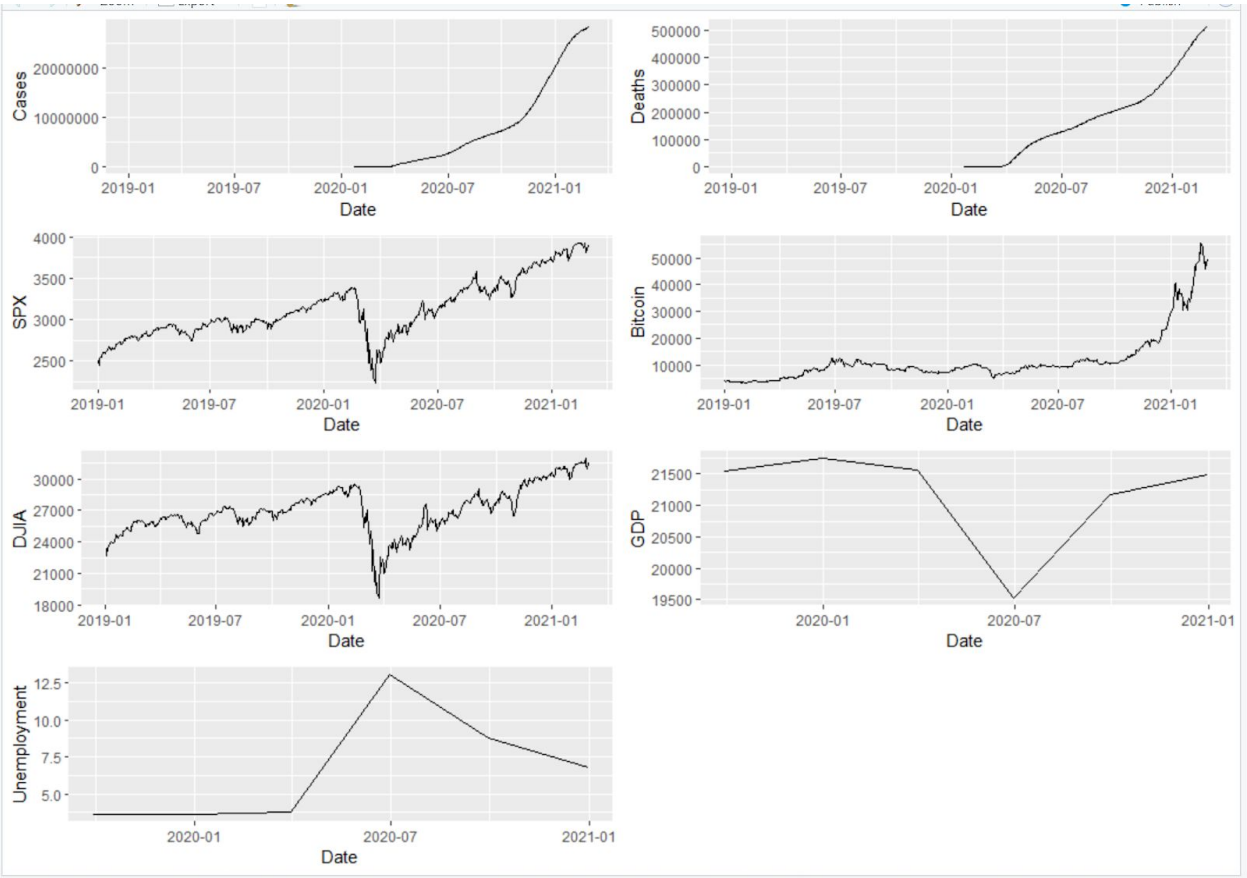
NDX, Gold and Oil Prices vs. Covid-19 Deaths



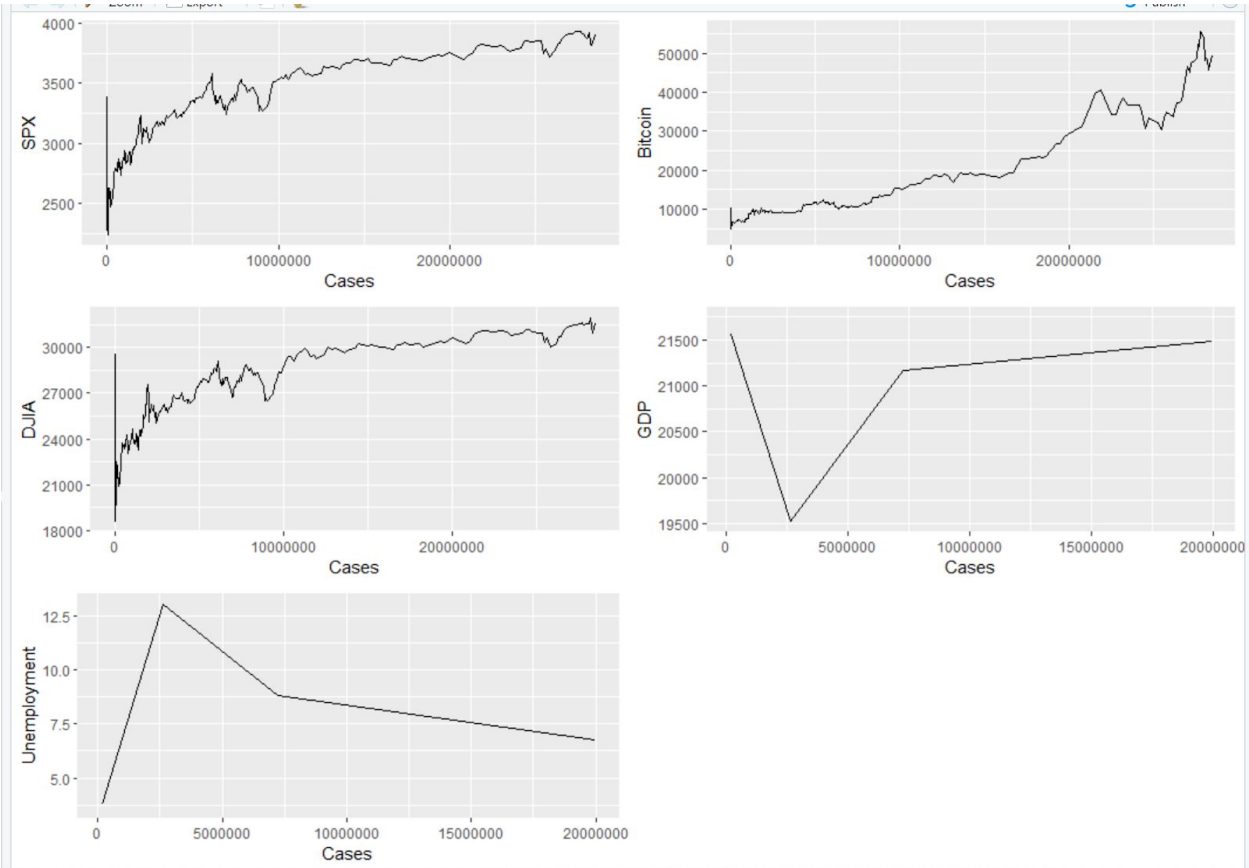
It can be seen from plots of prices vs. Cases or Death do not show any real correlation. One could make the assumption that the prices of NDX and Oil are positively correlated with both Cases and Deaths but that is more like just due to price increase over time, due to other factors.

There may be another factor at play, e.g. the initial news of the pandemic caused a scare that settled over time. Perhaps other factors such as macroeconomic indicators or the world's rapid improvement in virtualization provided stability to the market.

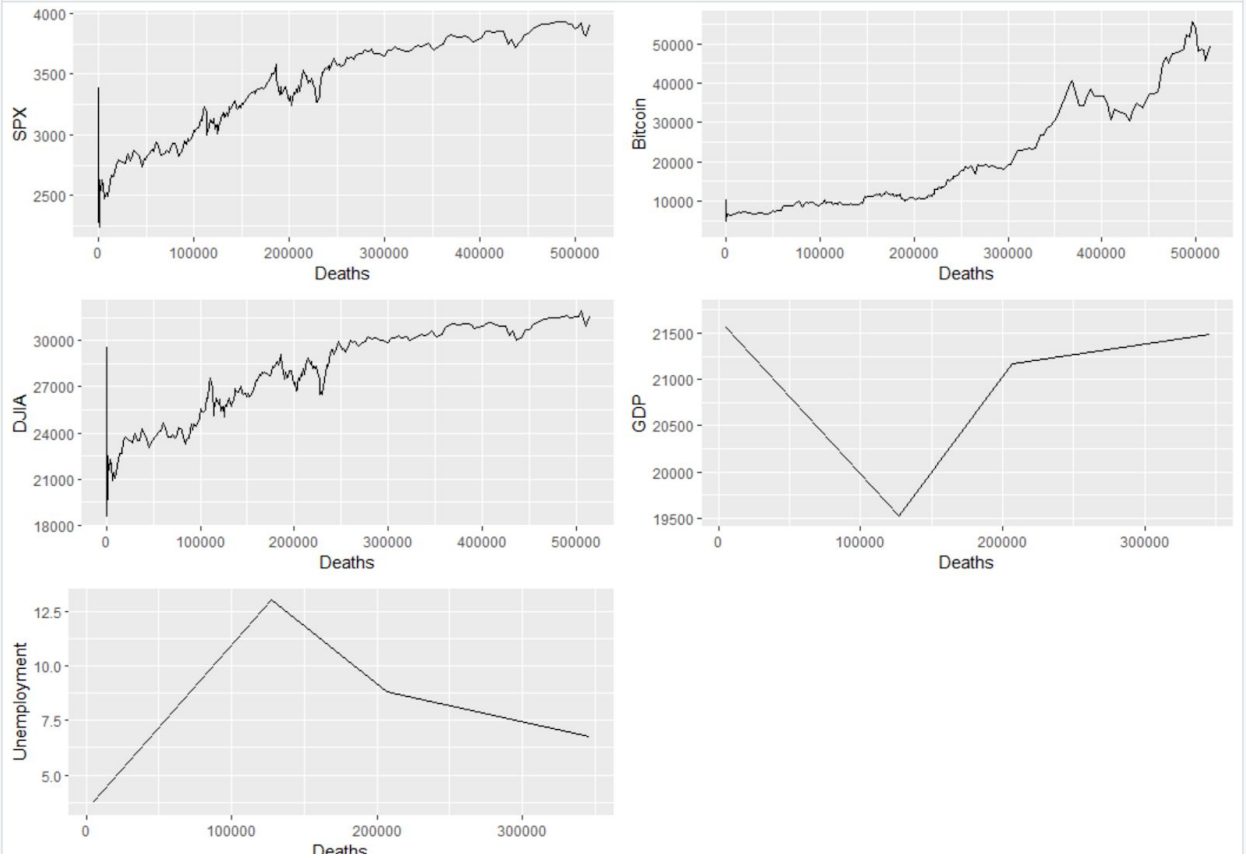
SPX, DOW, Bitcoin, GDP and Unemployment



From the plots above we can see that the SPX, Bitcoin, DJIA, GDP and Unemployment dropped significantly and then later rebounding in early 2021.



From the plots above we could see that the BTCUSD pair did not see significant fluctuation in its price during the pandemic, rest of the other markets responded to the case and their price were unstable until the cases hit 10M.



From the plots above we could see that the BTCUSD pair did not see significant fluctuation in its price during the pandemic, rest of the other markets responded to the case and their price were unstable until the death reached 100,000.

Methods:

Classification

1. Random forest
2. Regression
3. Wavelet transform

References:

[1] The relationship between cryptocurrencies and COVID-19 pandemic
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7388435/

[2] Impact of COVID-19 on forecasting stock prices: An Integration of Stationary Wavelet Transform and Bidirectional Long Short-Term Memory
https://www.hindawi.com/journals/complexity/2020/1846926/

[3] Bloomberg Terminal

Data	**Function**
TSA Throughput	TSATTPCY
CoVID-19 Cases	NCOVUSCA
CoVID-19 Recovered	NCOVUSRE
CoVID-19 Deaths	NCOVUSDE
CoVID-19 Vaccine administered	NCOVUSVA
S&P 500	SPX Index
Dow Jones	INDU Index
Nasdaq Composite	CCMP Index
Gold	XAU Curncy
Crude oil futures	C01 Comdty
Bitcoin/USD	XBTUSD
GDP, Non-farm payroll	ECST