For my final project, I mainly focused on the stock market and analyzed the market from 3 dimensions including: time series analysis, sentimental analysis, and excess return. I would illustrate each of them in the following passage.

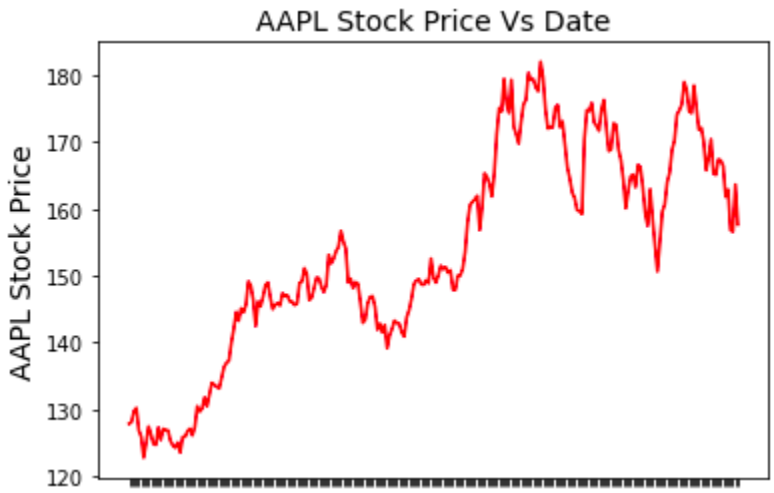
Motivation : I'm very interested in the stock market and have invested a bit in the tech company myself. And after a wonderful quarter of python learning. I wonder if there's anything I can do to better help people to make the investment. I'm aware that this has been a common object shared by many investors. But I still want to apply my knowledge on it.

Data Description: In this project, I have 3 csv files each containing the information about related news, stock historical prices, and prices of treasury bonds. I got the news from web scraping of finviz website, the stock price from API for stock data, and treasury bond price from Nasdaq API for national treasury data.

finviz website for scrapping

Time series analysis: The main objective of the analysis was to predict the future of a series of data based on a linear regression on the data itself. However, to make sure the model is effective, one has to make sure the data is stationary and is NOT a white noise data (a series of random numbers with mean as 0 and sd as 1, which can't be predicted).

I first plot the AAPL stock price vs data as the graph below.



After that, I'm trying to use the percent change of the stock price as the time series data to perform analysis, and the p_value of the Dickey Fuller Test is smaller than 0.05 so the data is stationary.

```
Results of Dickey-Fuller Test:
Test Statistic                -5.766704e+00
p-value                        5.511329e-07
#Lags Used                     7.000000e+00
Number of Observations Used    2.430000e+02
Critical Value (1%)           -3.457551e+00
Critical Value (5%)           -2.873509e+00
Critical Value (10%)          -2.573148e+00
dtype: float64
```

However, the p value of the White Noise Test is obviously larger than 0.05, so we can't reject the null hypothesis that this data series is not a White Noise Data. As a result, we have to change the data series to predict.

```
White Noise for pct change:      lb_stat  lb_pvalue
1  0.015967   0.899448
```

Then, I'm trying to use the close price as the time series data. Unfortunately, the data is not stationary.

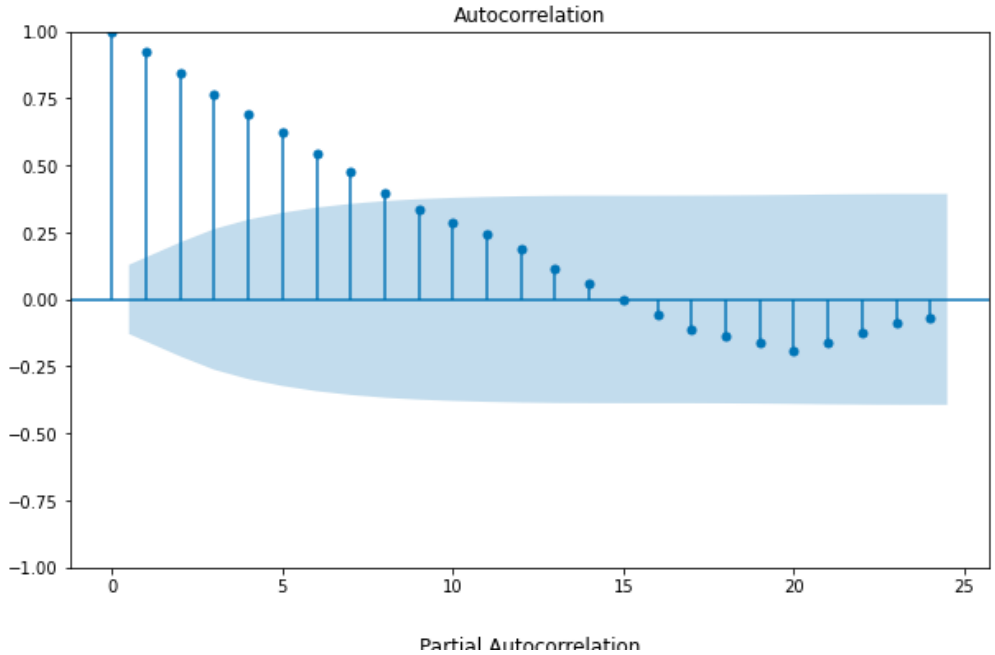
```
Results of Dickey-Fuller Test:
Test Statistic                  -2.156328
p-value                          0.222476
#Lags Used                       0.000000
Number of Observations Used    250.000000
Critical Value (1%)             -3.456781
Critical Value (5%)             -2.873172
Critical Value (10%)            -2.572969
dtype: float64
```

As a result, I take a difference of the stock price (which is using today's stock price subtract the stock price 20 days from now). And finally the data series is stationary now and is NOT a white noise data.

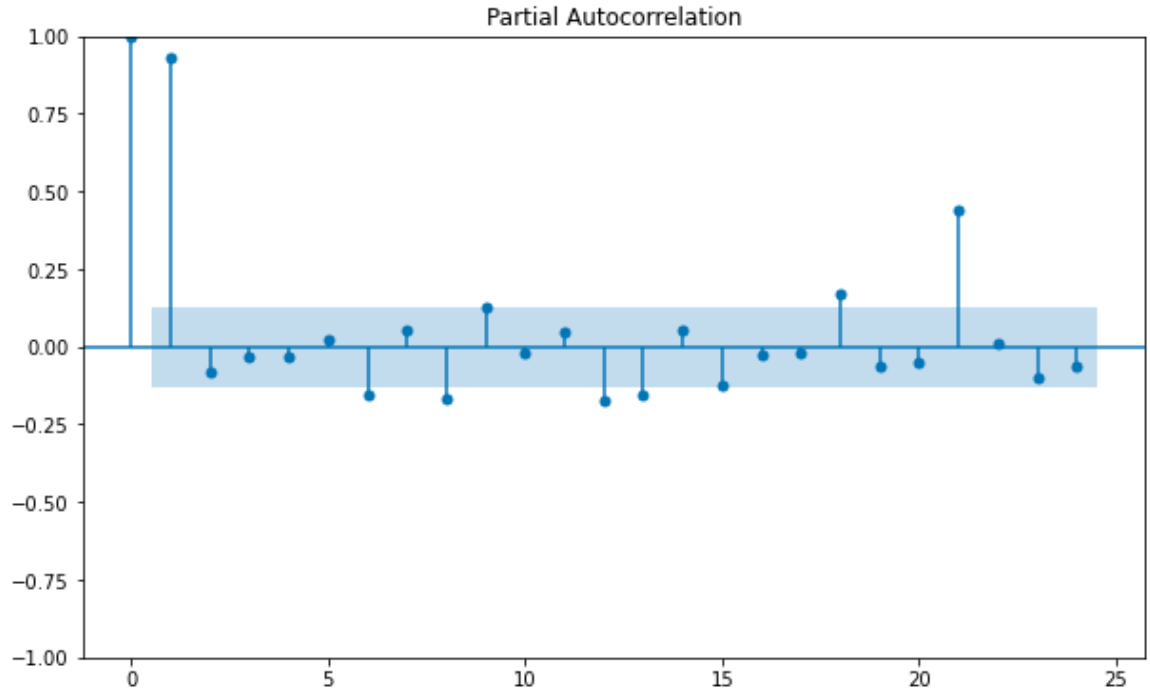
```
Results of Dickey-Fuller Test:
Test Statistic                  -3.728244
p-value                          0.003733
#Lags Used                      14.000000
Number of Observations Used    216.000000
Critical Value (1%)             -3.460992
Critical Value (5%)             -2.875016
Critical Value (10%)            -2.573952
dtype: float64
```

```
White Noise for difference:        lb_stat     lb_pvalue
1  201.813571  8.396122e-46
```

The next step is to choose which time series model to use (AR / MA / ARIMA). We determine this through the acf and pacf graphs.



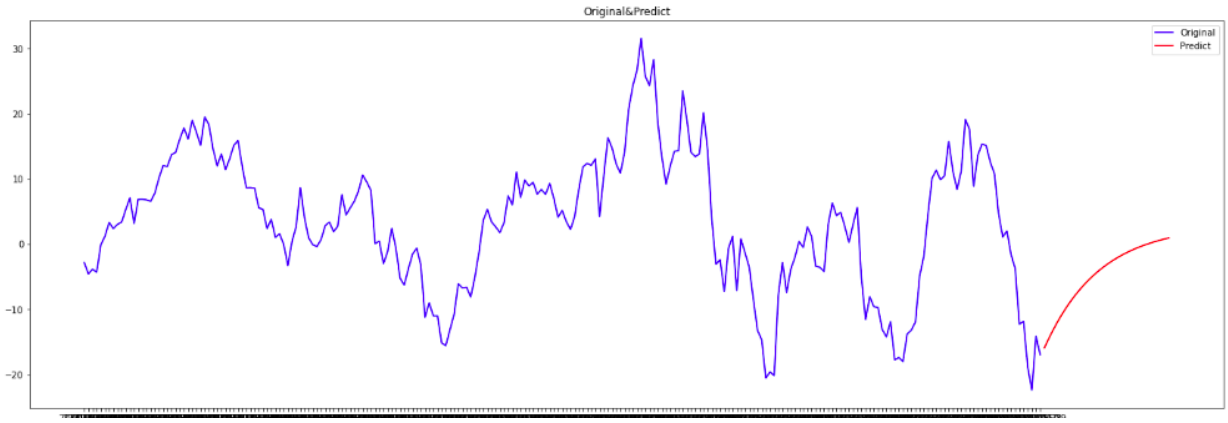
Partial Autocorrelation



Due to the ACF graph decreasing to 0 quickly and PACF becoming 0 suddenly after the second data, we use the AR model to forecast.

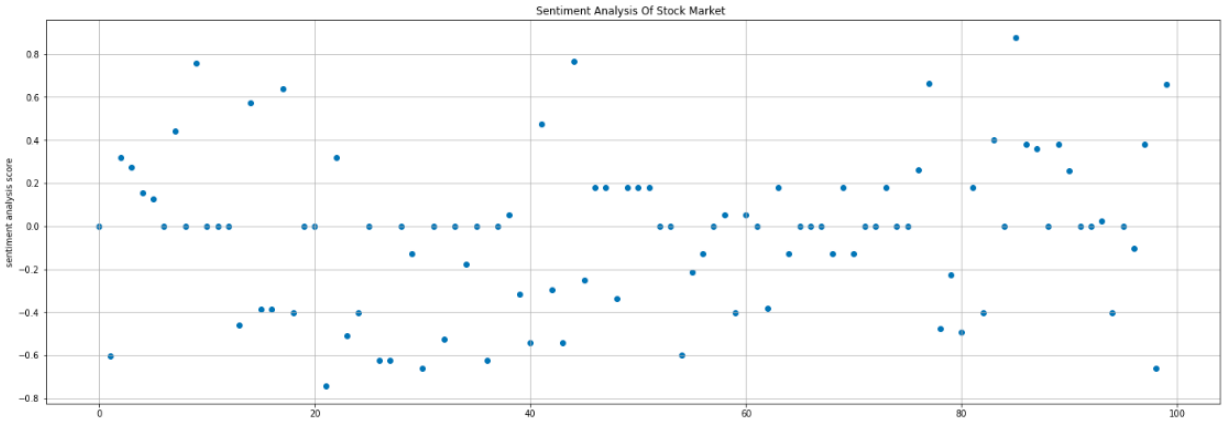
Then I used the dw index to evaluate the model, the mode is better when the value is close to 2. The dw value for this model is roughly 1.99

Finally, the prediction of my model is in the graph below.

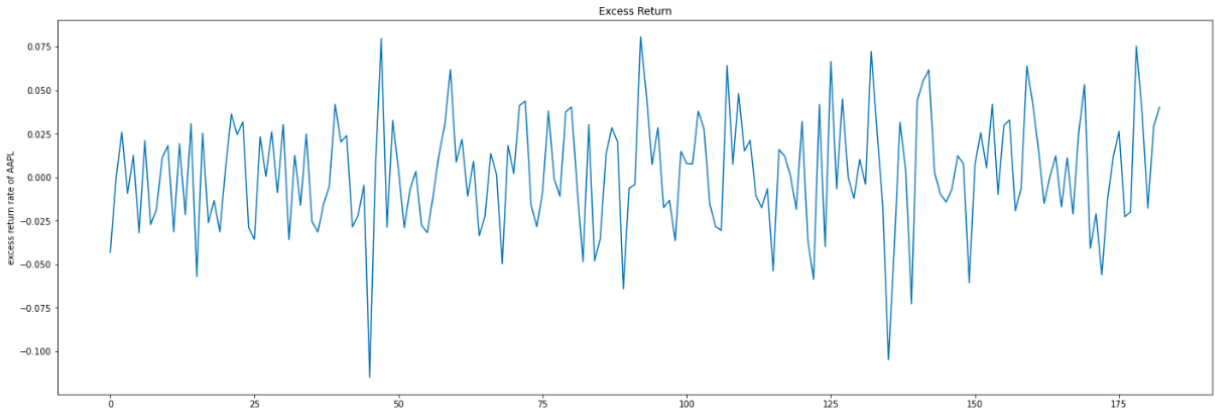


However, using time series to predict the stock market was HIGHLY unstable because the stock market can be affected by a lot of factors. One has to consider other information before he or she decides to invest in the stock.

Sentiment Analysis: The objective of the sentiment analysis is to evaluate the company based on the news. It would evaluate each news headline from bad(-1) to good(1). In this analysis, I used the SentimentIntensityAnalyzer() model which would return a polarity score of each news headline (Including a year range). I then plotted them in a dot graph and calculated the mean of it. The mean of it is roughly 0.023



Excess Return: The excess return is calculated by subtracting the percent change of the national treasury bond from the percent change of a particular stock price. Analysts would use this index to evaluate the performance of one stock (I used the treasury data from 2020-02-20. However due to the data of stock and the treasury not 100% coincide with each other, it only has 183 data left). The mean of the excess return is 0.002137468526829797.



Conclusion: The overall performance of the AAPL stock is very stable from the 3 analyses I have done. Nevertheless, The stock market is highly volatile and can be affected by many factors dramatically. So these information can only provide suggestions.