Project 2

Investment Analysis for Apple Stock

Group 4
Manasvi Mittal
Shrishail Kandi
Vikram Bhardwaj
Syed Hussain

Table of Content

Introduction

- We are looking from the perspective of an investment firm. Looking to explore opportunities to trade in the Apple stock ticker ($AAPL).
- Below we tried to find out any anomalies in the market the firm could explore to earn profits by trading.
- We have used tweets which contain the Apple stocks ticker for sentiment analysis of the market and stock price dataset from yahoo.

Data Processing and Cleaning

We have used 2 data sets for our project. The first dataset is the stocks data set, in which we are looking at Apple stock($AAPL). The Apple stocks data was taken from yahoo API. The other data set is the Stocks Tweet data that was taken from Kaggle and was cleaned according to our purpose. We followed these steps for discovering, structuring, cleaning and enriching the data in our data processing steps:

Discovering

The apple stocks data is from 2015 to 2020. The column names and the data types are below.

```
> colnames(apple_price)
[1] "Date"      "Open"      "High"      "Low"       "Close"     "Adj.Close" "Volume"
> sapply(apple_price, class)
       Date        Open        High         Low       Close   Adj.Close      Volume
"character"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"
```

For the tweets data, the column names are:

```
```{r}
colnames(tweets)
```

 [1] "tweet_id"   "writer"   "post_date"   "body"   "comment_num" "retweet_num" "like_num"
```

And the data type for these are "character"

Structuring

For the Apple stocks data, the date column was not in the standard format, so we had to change that to the standard format.

For the Tweets dataset, we selected the "body" of the data which consists of the tweets. As you can see, there is a lot of characters and data in the tweets that is not required, we clean it all in the next step. The body looks like this when it is uncleaned:

```
[1] "lx21 made $10,008  on $AAPL -Check it out! http://profit.ly/1MnD8s?aff=202 Learn #howtotrade http://bit.ly/1c1Nljx $EXE $WATT $IMRS
$CACH $GMO"
[2] "Insanity of today weirdo massive selling. $aapl bid up 45 cents after hours after non stop selling in trading hours"
[3] "S&P100 #Stocks Performance $HD $LOW $SBUX $TGT $DVN $IBM $AMZN $F $APA $GM $MS $HAL $DIS $MCD $BMY $XOM  more@
http://12Stocks.com/sp100"
[4] "$GM $TSLA: Volkswagen Pushes 2014 Record Recall Tally Higher https://pic.twitter.com/WIIc1lW7hW @ProTradersNews
http://growword.com/2015/01/01/0246.htmlâ€¦ @theferrarifan"
[5] "Swing Trading: Up To 8.91% Return In 14 Days http://ow.ly/GDks0 #swingtrading #forecast #techstock $MWW $AAPL $TSLA"
[6] "Swing Trading: Up To 8.91% Return In 14 Days http://ow.ly/GDkrL #swingtrading #forecast #techstock $MWW $AAPL $TSLA"
```

Cleaning

For the apple stocks dataset, we checked the null values in stock price and date and we found none of the null values.

```
> sum(is.na(apple_price$Close))
[1] 0
> sum(is.na(apple_price$Date))
[1] 0
```

We converted the date to a standard format which is more relevant to our analysis and later business questions.

For the tweets dataset, we had to remove all the unnecessary components in the data that would have resulted in an inaccurate result for clustering and the business questions. For our purpose, we are selecting and filtering the tweets that contain the ticker symbol "$AAPL" in the Tweet body. Since many tweets contain links to external articles, we removed all the "http" and "https" that might hamper the clustering process. After this, we converted all the characters in the tweet body to lower case and removed punctuation to maintain consistency. We also removed the non-alpha numeric values which were not useful to us.

Enriching

For the apple stocks data, we found the absolute returns and logarithmic return based on the price difference of a particular day and the day before it. The price difference was calculated on the basis of gain/loss percentage. The data we have is from 2015 to 2020, so we put them to various tests to determine which of the returns have the most amount of correlation with the past data to be used for time series regression.

```
> returns = 1- (apple_price_relevant/lag(apple_price_relevant, -1))
> returns
   2015-01-05    2015-01-06    2015-01-07    2015-01-08    2015-01-09    2015-01-12    2015-01-13    2015-01-14 
 2.817159e-02 -9.415529e-05 -1.402217e-02 -3.842227e-02 -1.072518e-03  2.464069e-02 -8.878719e-03  3.810524e-03 
   2015-01-15    2015-01-16    2015-01-20    2015-01-21    2015-01-22    2015-01-23    2015-01-26    2015-01-27 
 2.714029e-02  7.770118e-03 -2.575719e-02 -7.634327e-03 -2.601548e-02 -5.160178e-03 -1.062099e-03  3.501326e-02 
   2015-01-28    2015-01-29    2015-01-30    2015-02-02    2015-02-03    2015-02-04    2015-02-05    2015-02-06 
-5.653286e-02 -3.113350e-02  1.463411e-02 -1.254688e-02 -1.686251e-04 -7.669583e-03 -3.178388e-03  8.420910e-03 
   2015-02-09    2015-02-10    2015-02-11    2015-02-12    2015-02-13    2015-02-17    2015-02-18    2015-02-19 
-6.642563e-03 -1.921146e-02 -2.343878e-02 -1.265218e-02 -4.902736e-03 -5.901794e-03 -6.962372e-03  2.097607e-03 
   2015-02-20    2015-02-23    2015-02-24    2015-02-25    2015-02-26    2015-02-27    2015-03-02    2015-03-03 
-8.174418e-03 -2.702703e-02  6.240602e-03  2.557319e-02 -1.265633e-02  1.502831e-02 -4.904157e-03 -2.091595e-03 
   2015-03-04    2015-03-05    2015-03-06    2015-03-09    2015-03-10    2015-03-11    2015-03-12    2015-03-13 
 6.338961e-03  1.657062e-02 -1.503014e-03 -4.265403e-03  2.068583e-02  1.823153e-02 -1.807919e-02  6.910406e-03 
   2015-03-16    2015-03-17    2015-03-18    2015-03-19    2015-03-20    2015-03-23    2015-03-24    2015-03-25 
-1.100413e-02 -1.672672e-02 -1.125630e-02  7.550401e-03  1.254902e-02 -1.040508e-02  4.087698e-03  2.612683e-02 
   2015-03-26    2015-03-27    2015-03-30    2015-03-31    2015-04-01    2015-04-02    2015-04-06    2015-04-07 
-6.970336e-03  7.968416e-03 -2.531443e-02  1.535178e-02  1.446596e-03 -8.611670e-03 -1.619853e-02  1.052215e-02 
   2015-04-08    2015-04-09    2015-04-10    2015-04-13    2015-04-14    2015-04-15    2015-04-16    2015-04-17 
 3.253742e-03 -7.643280e-03 -4.266783e-03  1.966955e-03  4.335798e-03 -3.800443e-03  4.811484e-03  1.125466e-02 
   2015-04-20    2015-04-21    2015-04-22    2015-04-23    2015-04-24    2015-04-27    2015-04-28    2015-04-29 
-2.284569e-02  5.407492e-03 -1.347405e-02 -8.163614e-03 -4.704249e-03 -1.819153e-02  1.575572e-02  1.470585e-02 
   2015-04-30    2015-05-01    2015-05-04    2015-05-05    2015-05-06    2015-05-07    2015-05-08    2015-05-11 
 2.712998e-02 -3.036353e-02  1.938736e-03  2.253296e-02  6.279809e-03 -1.999840e-03 -1.884081e-02  1.018652e-02 
```

```
> returns_log = log(apple_price_relevant/lag(apple_price_relevant, -1))
> returns_log
   2015-01-05    2015-01-06    2015-01-07    2015-01-08    2015-01-09    2015-01-12    2015-01-13    2015-01-14 
-2.857602e-02  9.415086e-05  1.392477e-02  3.770252e-02  1.071943e-03 -2.494935e-02  8.839534e-03 -3.817803e-03 
   2015-01-15    2015-01-16    2015-01-20    2015-01-21    2015-01-22    2015-01-23    2015-01-26    2015-01-27 
-2.751539e-02 -7.800463e-03  2.543106e-02  7.605333e-03  2.568283e-02  5.146910e-03  1.061536e-03 -3.564092e-02 
   2015-01-28    2015-01-29    2015-01-30    2015-02-02    2015-02-03    2015-02-04    2015-02-05    2015-02-06 
 5.499266e-02  3.065868e-02 -1.474225e-02  1.246882e-02  1.686109e-04  7.640321e-03  3.173347e-03 -8.456566e-03 
   2015-02-09    2015-02-10    2015-02-11    2015-02-12    2015-02-13    2015-02-17    2015-02-18    2015-02-19 
 6.620598e-03  1.902925e-02  2.316831e-02  1.257281e-02  4.890757e-03  5.884447e-03  6.938247e-03 -2.099810e-03 
   2015-02-20    2015-02-23    2015-02-24    2015-02-25    2015-02-26    2015-02-27    2015-03-02    2015-03-03 
 8.141189e-03  2.666825e-02 -6.260155e-03 -2.590586e-02  1.257690e-02 -1.514238e-02  4.892170e-03  2.089411e-03 
   2015-03-04    2015-03-05    2015-03-06    2015-03-09    2015-03-10    2015-03-11    2015-03-12    2015-03-13 
-6.359138e-03 -1.670945e-02  1.501886e-03  4.256332e-03 -2.090278e-02 -1.839977e-02  1.791770e-02 -6.934393e-03 
   2015-03-16    2015-03-17    2015-03-18    2015-03-19    2015-03-20    2015-03-23    2015-03-24    2015-03-25 
 1.094402e-02  1.658837e-02  1.119342e-02 -7.579049e-03 -1.262842e-02  1.035132e-02 -4.096075e-03 -2.647420e-02 
   2015-03-26    2015-03-27    2015-03-30    2015-03-31    2015-04-01    2015-04-02    2015-04-06    2015-04-07 
 6.946155e-03 -8.000334e-03  2.499933e-02 -1.547083e-02 -1.447644e-03  8.574801e-03  1.606874e-02 -1.057790e-02 
   2015-04-08    2015-04-09    2015-04-10    2015-04-13    2015-04-14    2015-04-15    2015-04-16    2015-04-17 
-3.259047e-03  7.614218e-03  4.257706e-03 -1.968892e-03 -4.345225e-03  3.793240e-03 -4.823097e-03 -1.131847e-02 
   2015-04-20    2015-04-21    2015-04-22    2015-04-23    2015-04-24    2015-04-27    2015-04-28    2015-04-29 
 2.258864e-02 -5.422166e-03  1.338408e-02  8.130472e-03  4.693219e-03  1.802804e-02 -1.588116e-02 -1.481506e-02 
   2015-04-30    2015-05-01    2015-05-04    2015-05-05    2015-05-06    2015-05-07    2015-05-08    2015-05-11 
-2.750479e-02  2.991168e-02 -1.940618e-03 -2.279071e-02 -6.299610e-03  1.997843e-03  1.866552e-02 -1.023876e-02 
   2015-05-12    2015-05-13    2015-05-14    2015-05-15    2015-05-18    2015-05-19    2015-05-20    2015-05-21 
-3.568710e-03  1.111641e-03  2.306340e-02 -1.396803e-03  1.096705e-02 -9.221241e-04 -7.697690e-05  1.017415e-02 
   2015-05-22    2015-05-26    2015-05-27    2015-05-28    2015-05-29    2015-06-01    2015-06-02    2015-06-03 
 8.714427e-03 -2.227736e-02  1.849778e-02 -1.970981e-03 -1.144789e-02  1.993652e-03 -4.452860e-03  1.230299e-03 
```

By using Ljung test, we found that the log returns had a higher correlation with the past data. We also used generalized autoregressive conditional heteroskedasticity (GARCH) model in our business questions to predict the volatility in the stock price.

For the tweets data, we had to remove the stop words such as "the", "of", "to", "in", etc. This would help us create a more accurate word cloud and clustering for our business questions. The final cleaned data tweets looks like this:

	unique.aapl_tweets.body. <chr>
1	lx21 made 10008 aapl check
2	swing trading 891 return 14 days
3	swing trading 891 return 14 days
4	swing trading 891 return 14 days
5	swing trading 891 return 14 days
6	swing trading 891 return 14 days

6 rows

Business Problems

Here we take a look at the business questions from the perspective of an investment firm looking to invest in the stock of Apple Inc. Here we cover diverse questions based on stock market like how investors can understand the price trends of AAPL based on the stock price over the years.

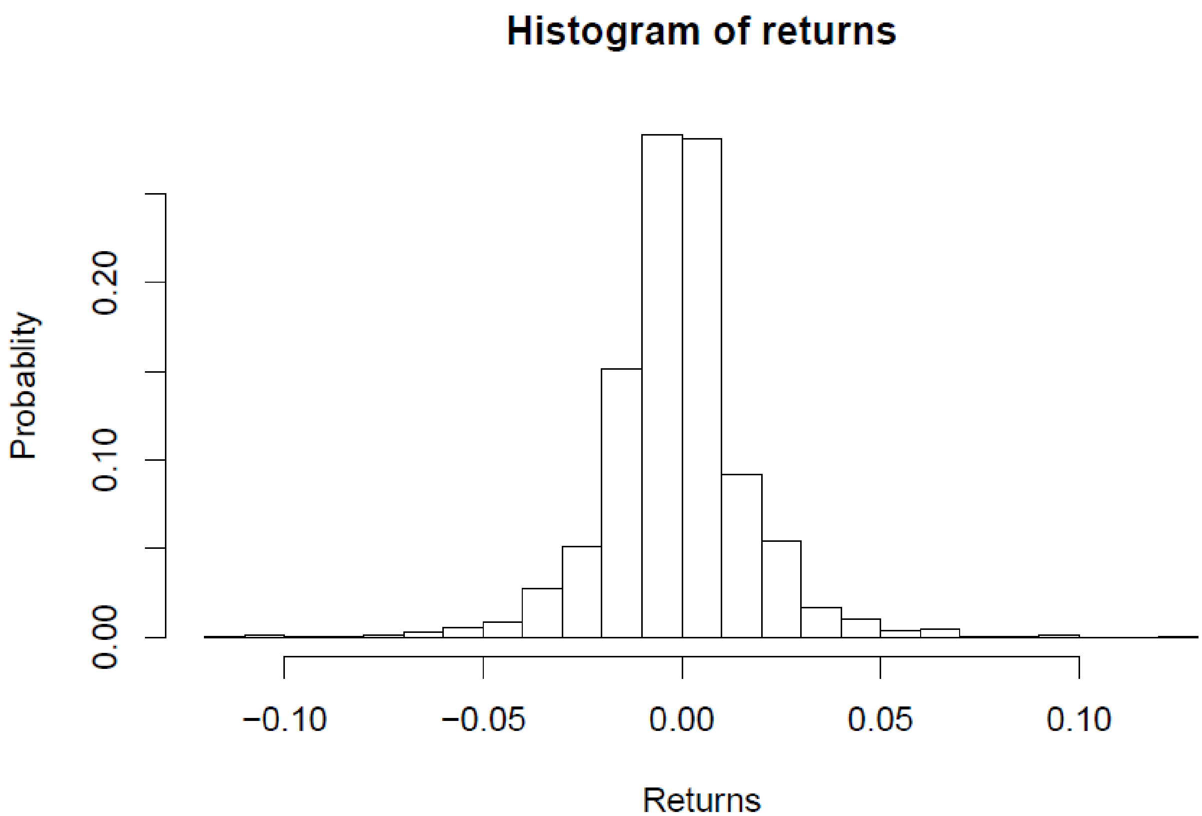
Based on the stock price data obtained from the previous year's how we can plan what the best way to invest in volatile markets is and based on it we can find the model to predict the volatility of the stock price to build our investment models.

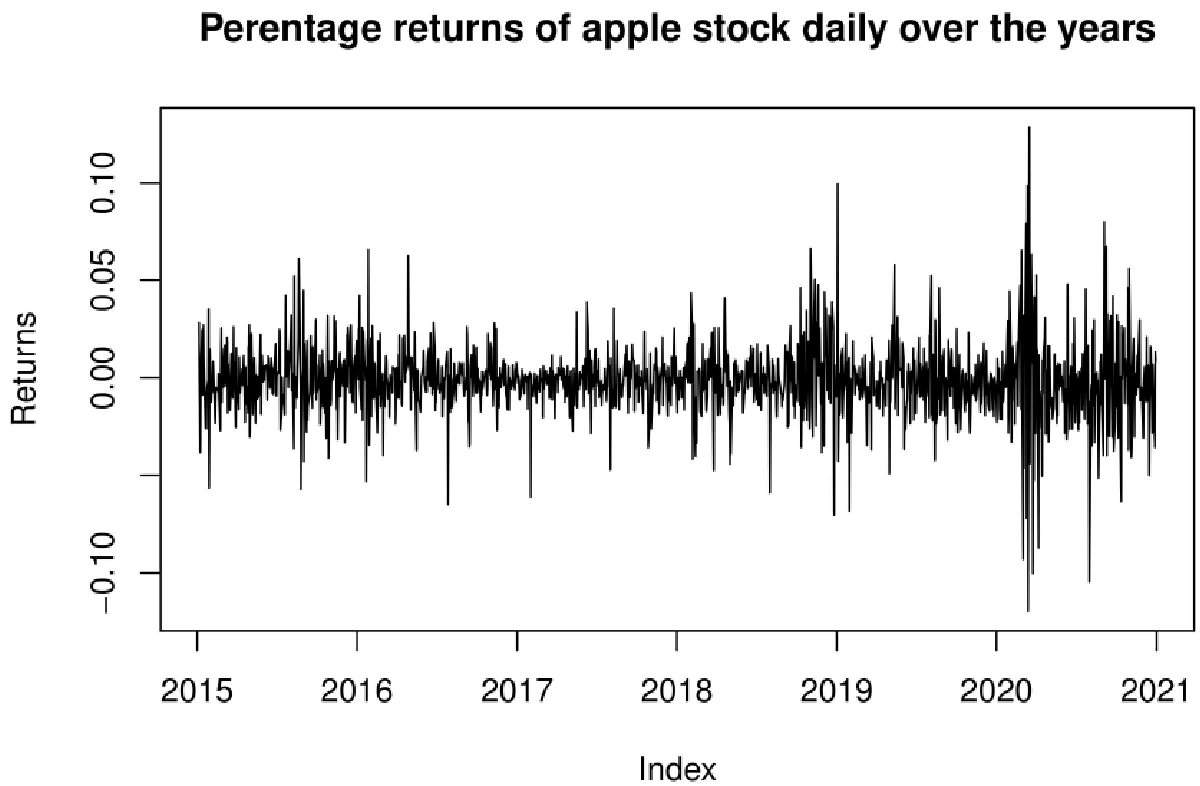
1. Let's understand the previous price trends of the Apple inc. ticker?

As you can see from the representation below, the apple stock price has historically compounded over the years. Historically, apple stock has given a compounded annual return of around 33%. The stock had fallen recently due to the corona crisis to $55 in March of 2019 and recovered to the new high levels of 130's by December of 2020. This is the summary of the Apple stock price over the years.

2. How have the daily returns of Apple stock been between 2015- 2020?

The graph below gives us the probability distribution of daily returns given by Apple Inc stock between the years 2015 and 2020. Above distribution shows that that Apple stock is not a highly volatile stock with almost 50 percent of all the daily returns lying in between -1% / +1% range.



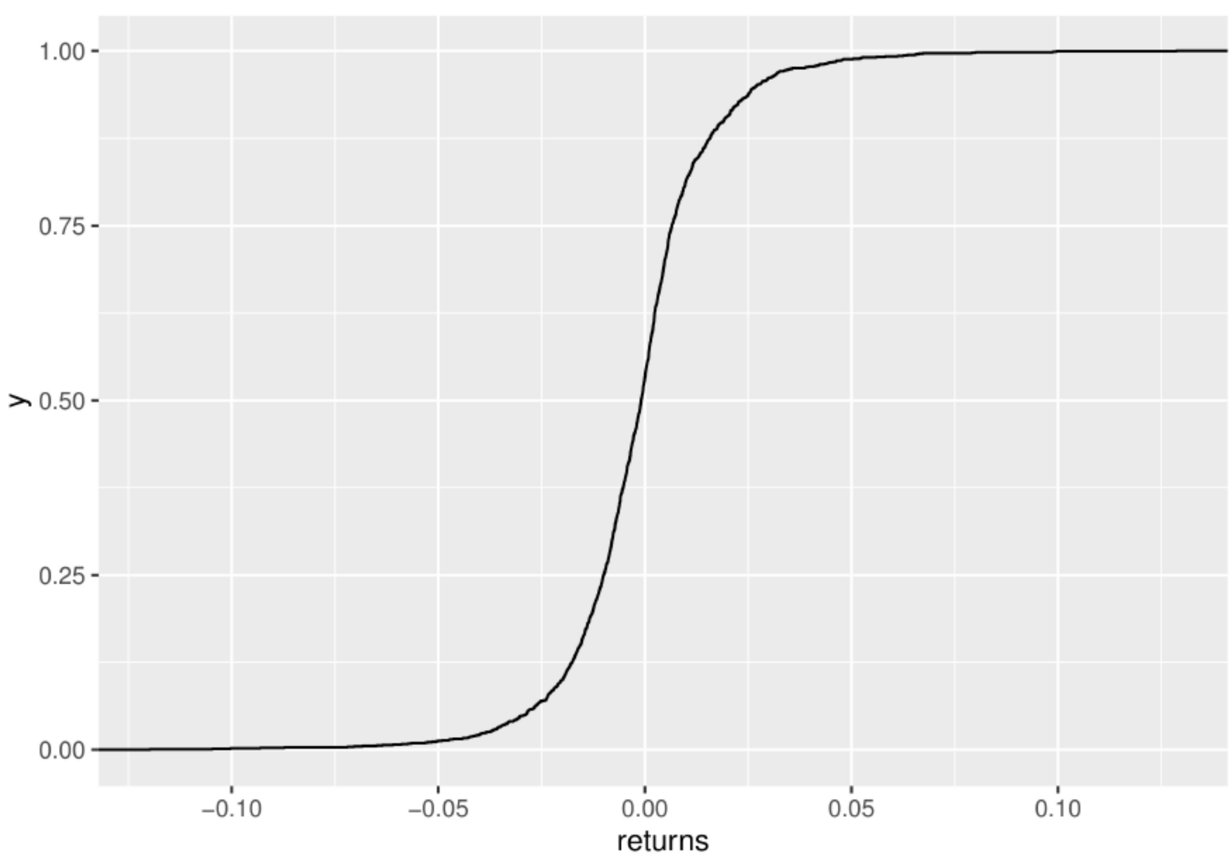

Perentage returns of apple stock daily over the years
Returns
0.10
0.05
0.00
−0.10
2015
2016
2017
2018
2019
2020
2021
Index

3. Can a long/short position in apple stock generate more than a 5% return in a day (considering absolute returns)?

As you can see for the above return cdf and percentile data, historically about 1.25% of the times the stock has fallen more than 5% between adjacent days closing and about 1% of the time the stock price has increased more than 5% between adjacent days closing. This shows us that the probability of apple stock generating returns greater than 5 percent given a long/short position (if returns are in absolute investment amounts without leverage) is less than 2.5%. This means the apple stock barely reacts drastically to the market conditions/news and hence you have a very less probability of making more than five percent in a given day, by trading Apple stock.

```
##         1.25%          2.5%         3.75%            5%         6.25%
## -0.0492362924 -0.0377879885 -0.0332398208 -0.0288431718 -0.0264176087
##          7.5%         8.75%           10%        11.25%         12.5%
## -0.0235913496 -0.0219014575 -0.0198662718 -0.0188620509 -0.0175513902
##        13.75%           15%        16.25%         17.5%        18.75%
## -0.0166471859 -0.0155056042 -0.0148740021 -0.0140258121 -0.0133491657
##           20%        21.25%         22.5%        23.75%           25%
## -0.0125170434 -0.0119165864 -0.0111387788 -0.0104965623 -0.0100170492
##        26.25%         27.5%        28.75%           30%        31.25%
## -0.0093747668 -0.0087408877 -0.0083075705 -0.0079255993 -0.0074619544
##         32.5%        33.75%           35%        36.25%         37.5%
```

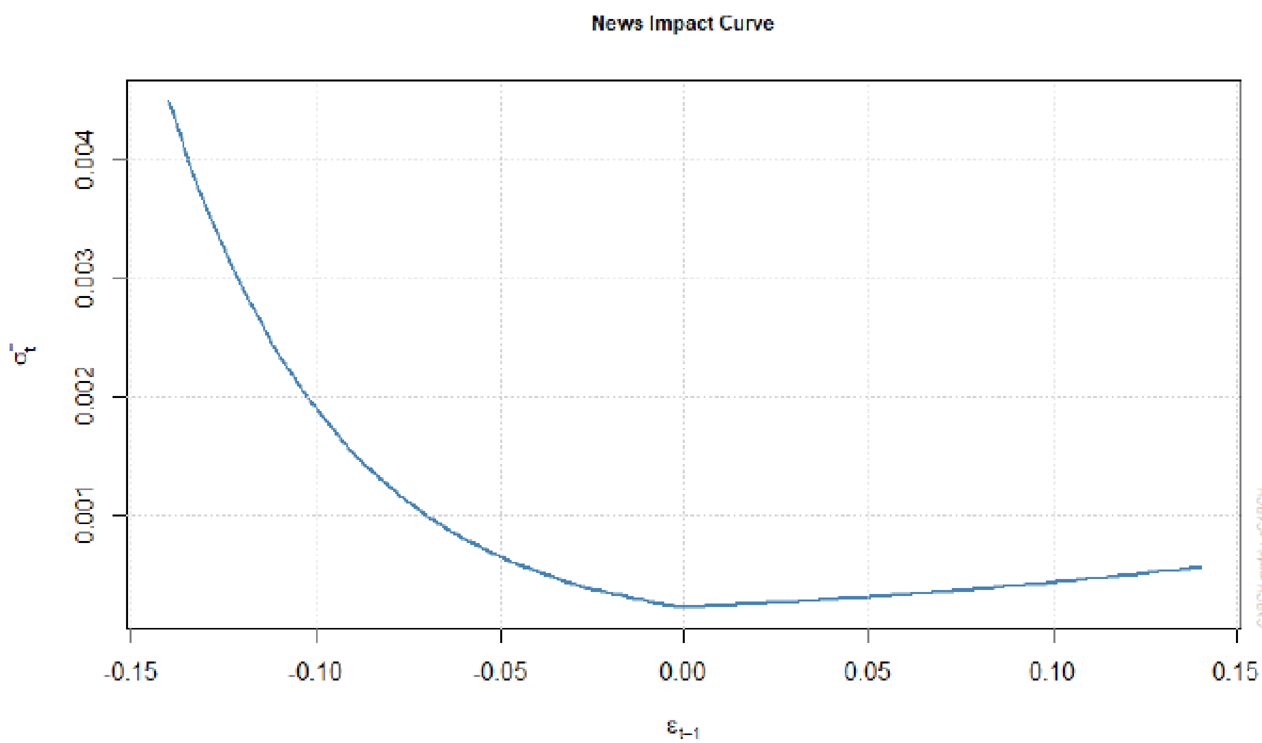
```
## -0.0070806177 -0.0065900265 -0.0061312162 -0.0058919212 -0.0052639969
##        38.75%           40%        41.25%         42.5%        43.75%
## -0.0047963007 -0.0042986313 -0.0037996921 -0.0034621941 -0.0030989284
##           45%        46.25%         47.5%        48.75%           50%
## -0.0026749905 -0.0020463813 -0.0016500715 -0.0011967839 -0.0008932886
##        51.25%         52.5%        53.75%           55%        56.25%
## -0.0004926635 -0.0002219613  0.0000941117  0.0004398384  0.0008832756
##         57.5%        58.75%           60%        61.25%         62.5%
##  0.0011195653  0.0014489716  0.0018734563  0.0022001140  0.0023415033
##        63.75%           65%        66.25%         67.5%        68.75%
##  0.0028653531  0.0031934879  0.0036842991  0.0041338879  0.0044577835
##           70%        71.25%         72.5%        73.75%           75%
##  0.0047719982  0.0052949811  0.0056349215  0.0059241400  0.0065276879
##        76.25%         77.5%        78.75%           80%        81.25%
##  0.0072512627  0.0077623130  0.0084278646  0.0092658455  0.0098685630
##         82.5%        83.75%           85%        86.25%         87.5%
##  0.0108151029  0.0115743858  0.0131148413  0.0144946597  0.0157330804
##        88.75%           90%        91.25%         92.5%        93.75%
##  0.0171054329  0.0190040821  0.0207088545  0.0225979790  0.0250405942
##           95%        96.25%         97.5%        98.75%          100%
##  0.0270243351  0.0306380897  0.0356442816  0.0478127973  0.1286469475
```



Time Series Analysis

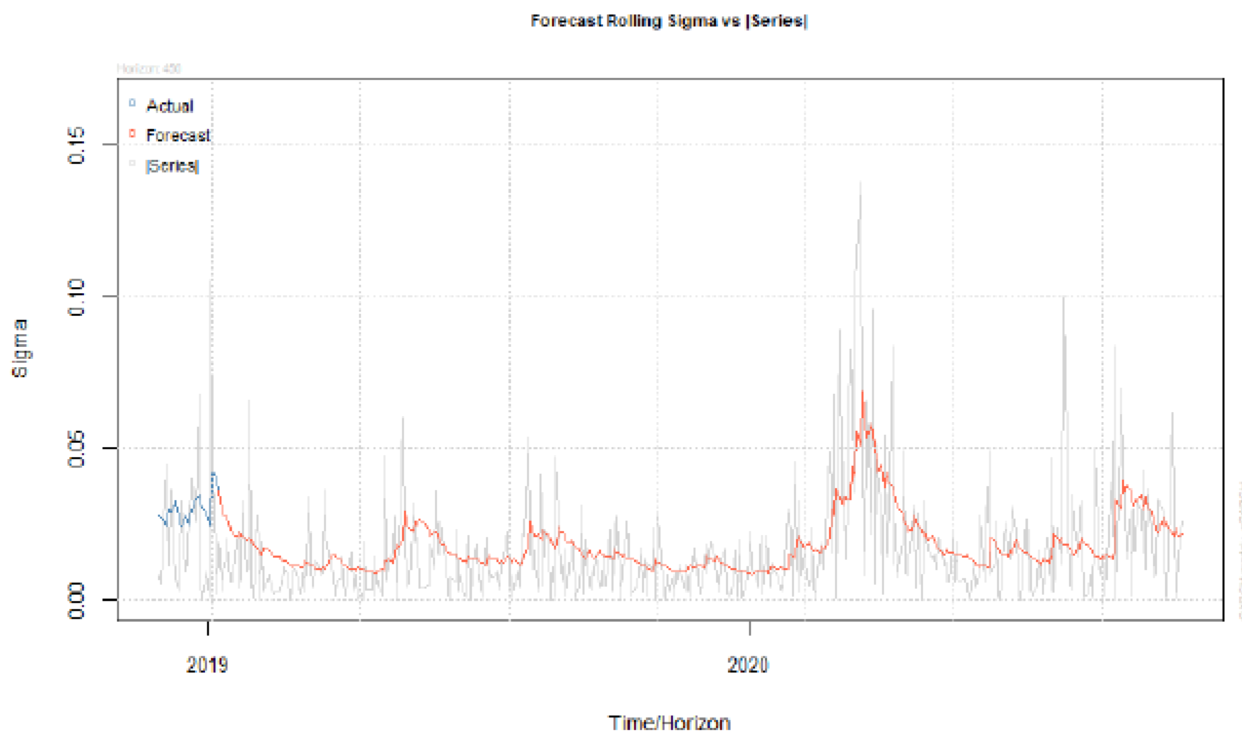
4. Can we make a strategy based on the volatility of the stock based on previous stock price data?

While trying to find out the variation of price change percentage while performing time series prediction using GARCH model. We found out that the volatility of percentage price change is much higher when the price is falling as compared to the volatility of percentage price change when the price is increasing. Using this nature of the ticker we can use strategies which make profit based on the volatility of the stock price. The trader goes both long and short on a stock at the same time creating a band on both the positive and negative side of the current price within which the trader is at loss. Whenever the ticker exceeds the price, the trader will earn a profit. This News Impact curve helps us understand that this can likely be done in case of the apple stock when the price starts falling as the price volatility of the stock drastically increases.



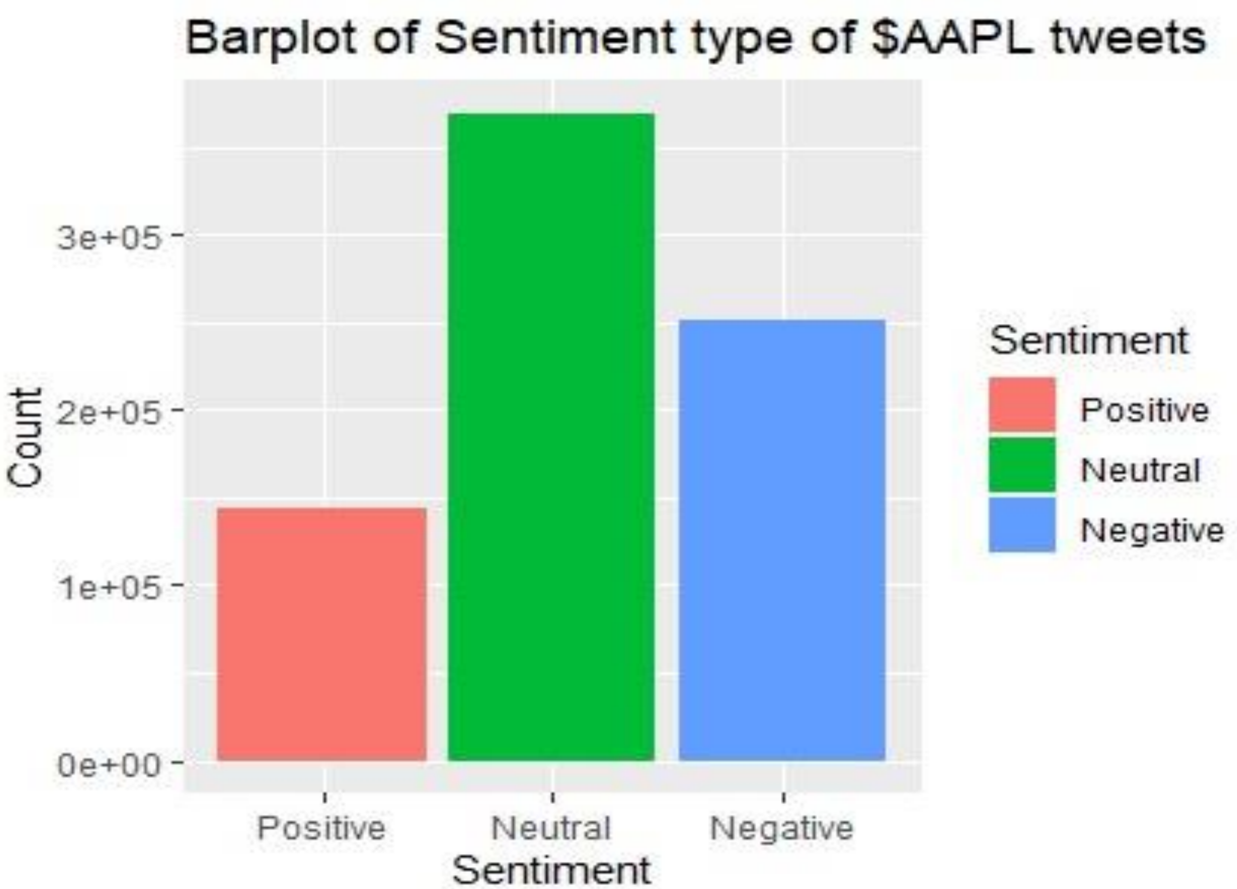
5. Continuing on the above question can we find the model to predict the volatility of the stock price to build our investment models?

We were able to adequately predict the volatility or the sigma of the stock price with adequate effect. This model can be fine-tuned with better data sets of stock prices with each transaction being reflected in the data set instead of just daily data. These kinds of datasets are hard to access and need to be bought from brokers on whose exchange the transactions are taking place and hence, they are very expensive. The regression model looks adequate in predicting volatility given our data sets.



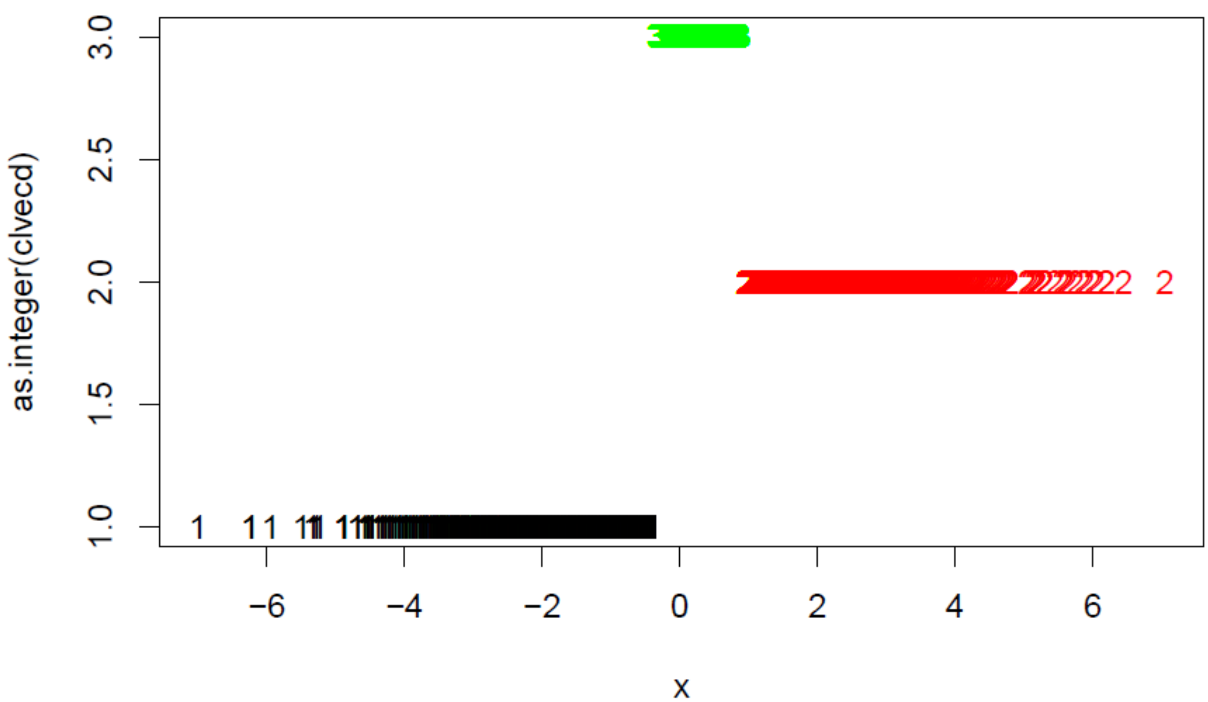
6. Is the market price of Apple stock related to the sentiment of that stock amongst investors? (We will use twitter's tweet data to find out the sentiment of the stock amongst investors)

As shown by our results for all the years below. We can see that the sentiment for the apple stock does not reflect the apple stock price. The stock price of apple has increased more than three hundred percent over the years whereas the general sentiment of apple stock over the years has been largely negative or neutral based on our tweets data. Let us further study the relationship of market sentiment and share price to further establish the relationship between the both of them using as market events in the next question.



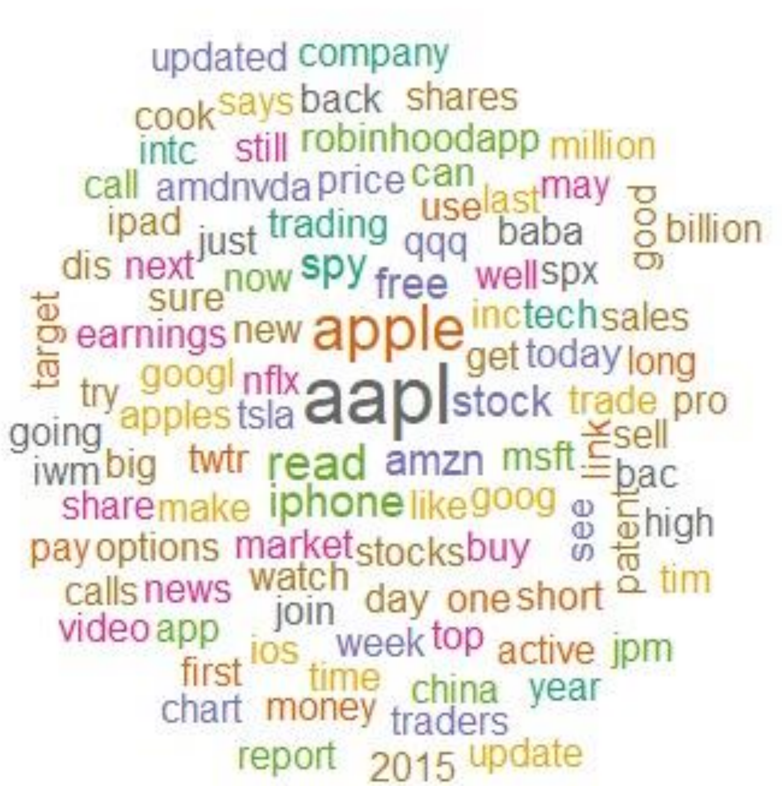
7. What was the general sentiment in October of 2018, which started a huge market collapse?

It can be noted that the clustering of tweets yields the following results: For the month of October 2018, our clustering model shows that the sentiment of the market was largely negative to neutral with over 66% of tweets in this range. The sentiment analysis model can be further used to make an investment strategy as our model's results for the market sentiment correlates with $AAPL's stock price during this time period. It can also be noted that the variance of the positive tweets is more when compared to the other two.



8. Word cloud of all the different words and their weightage in the entire corpus

As we can see in the image, Apple's products such as iPhone, iPad are talked about at large. People also talk about the effect China has on Apple since they are known to hire a lot of cheap labor from China to assemble their products. Also, we can conclude from the word cloud that the general consensus is to compare Apple's stock to other tech giants like Amazon (AMZN), Google (GOOG) and Microsoft ($MSFT).



Conclusion

There have been few Apple product releases that immediately resulted in a meteoric rise in the company's stock price. Day traders are known to target Apple at the release of each of its products, but the quick riches that they seek are all too often a mirage that swiftly disappears.

On the other hand, all their products had a positive effect on the stock over a longer period of time. The overarching, long-term view is the one to properly frame your investment decisions on, not day-to-day volatility. Over time, the market mechanism will identify true value in the marketplace. Rely on the wisdom of the masses over the long term, not on the speculators that routinely come and go, thereby letting companies like Apple work for you.

As of December 2021, Apple has a market cap of $2.944 Trillion. This makes Apple the world’s most valuable company by market cap.