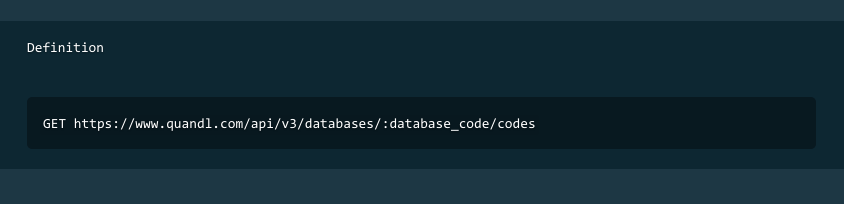
**Quandl database FinSentS Web News Sentiment vs Selected S&P 500 Stock Price Analysis**

Aim

Do a concise analysis about Trisent sentiment analysis result by comparing to market performance.

Sentiment and Market Data



The data provided by FinSents include the stock News Buzz, News Volume and Sentiment value ranging from -5 to 5, with 0 as no data available. However, due to the limited calls for free account, I had hand the stock market data from Google finance shown in the below table.

|  |  |
| --- | --- |
| **Industry Sectors** | **Company** |
| Car Manufacturing | Tesla, General Motor |
| Manufacturing | 3M |
| Construction Machinery and Equipment | Catepillar Inc |
| IT | Apple, Facebook, Amazon, Microsoft, Ebay, Yahoo |
| Oil and Energy | Exxon Mobil |
| Electronics | General Electrics, Intel Corp |
| Agriculture | Syngenta, Monsanto |
| Pharmacy | Allergan Inc, Abbott Laboratories |
| Banking and Financial | JP Morgan, Bank of America, Master Card |

These stocks are leading companies in their representing industries ranging from agriculture to IT sector.

For the ease of future analysis, I parsed data as python pandas dataframe structure and saved them into pickle module.

Analysis

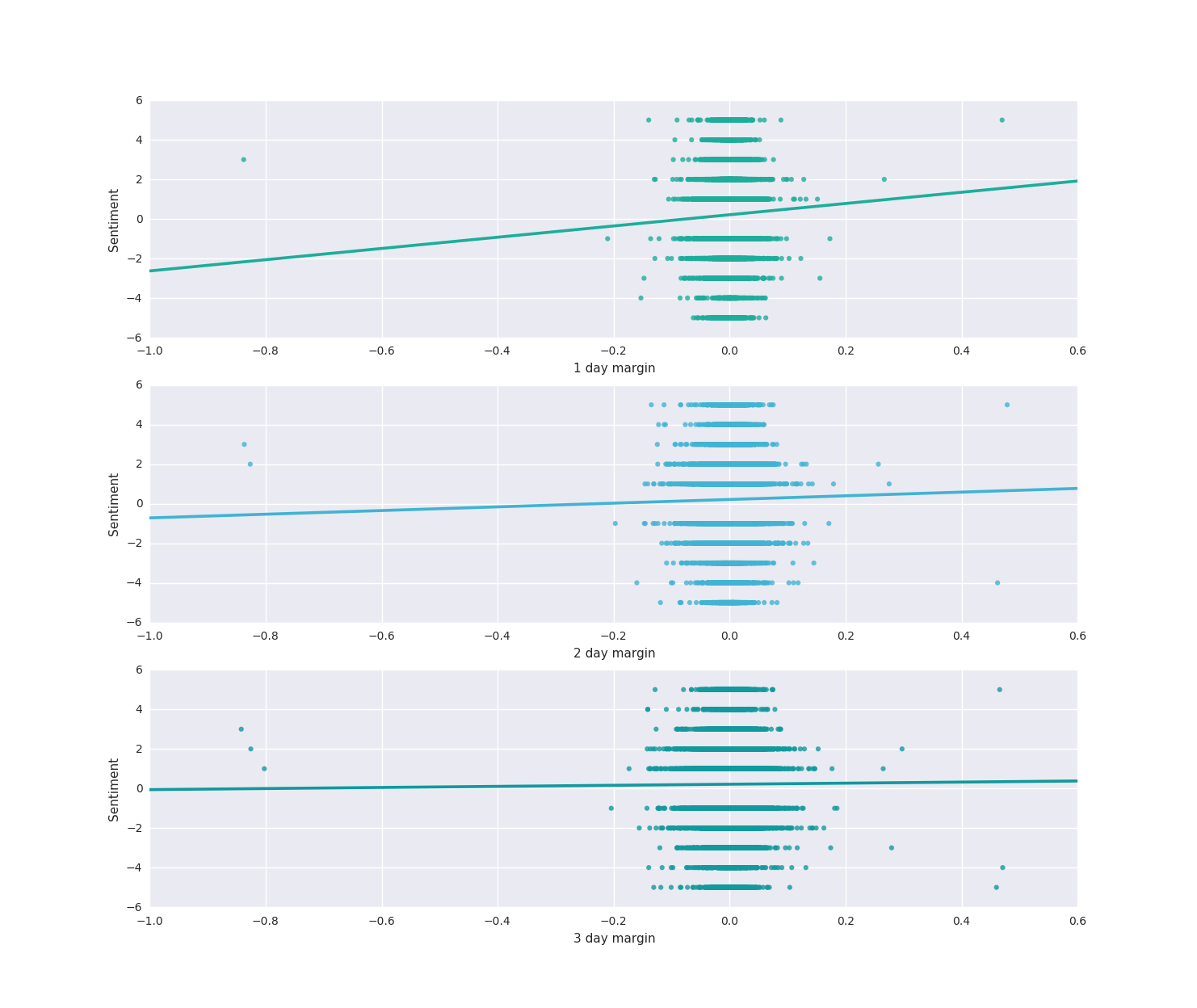
To extract the full correlation of Finsents sentiment data, I had compare to the sentiment, news buzz and news volume with future and past log return value. The results of analysis are visualized using python seaborn and matplotlib plotting library.

As for the market data, I extracted the selected raw data from Quandl Google finance database. The raw data is then sorted into day 1 to day 7, with each day corresponds to the day market closing price. Hence, I will be able to compare the returns from each day, if I had bought the stock at the day opening price and sold it after holding a period of days.

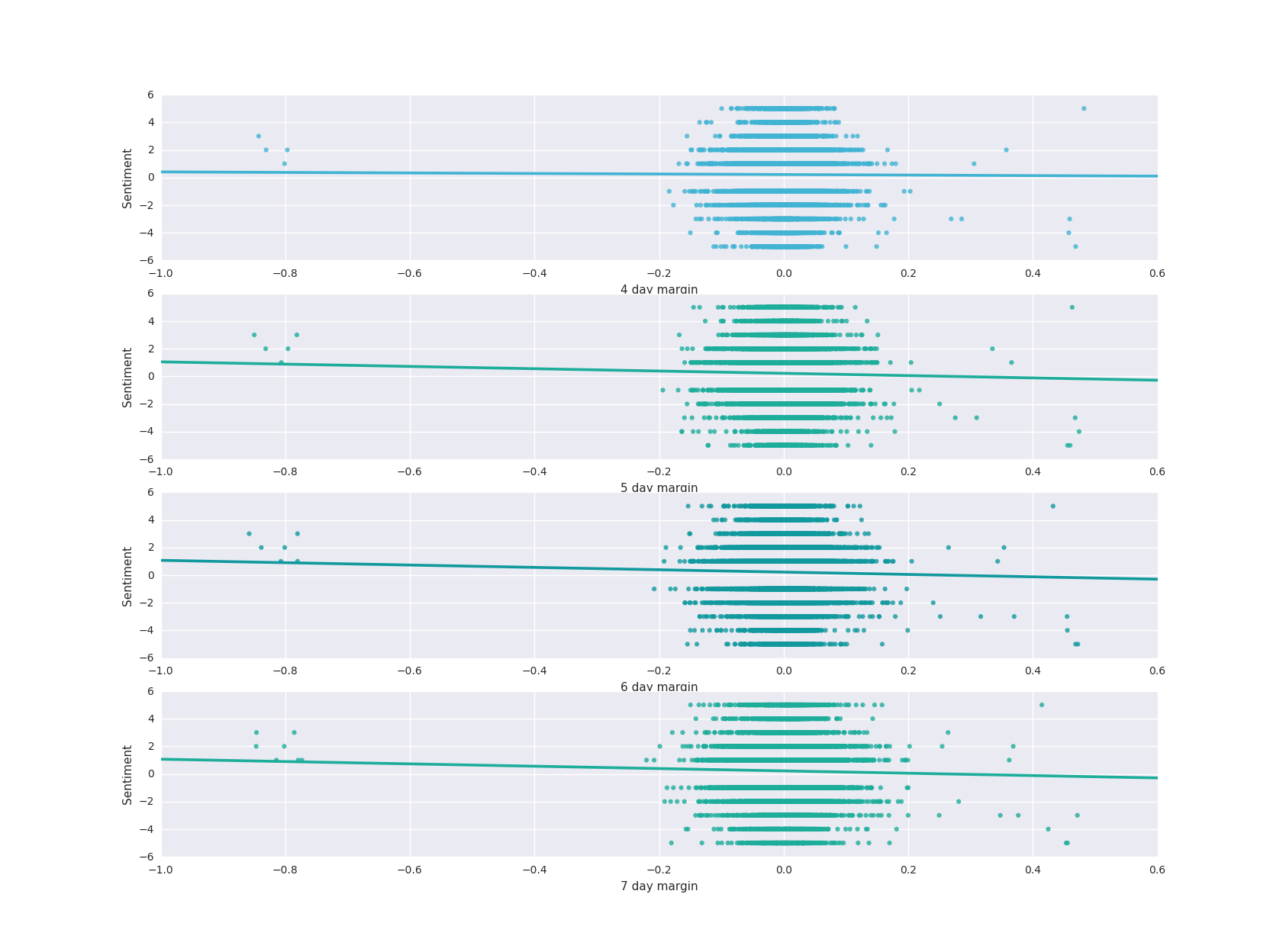
On the other hand, the closing price of the past day 1 till day 7 is collected for later comparison. These data are extracted and stored as dictionary structure and serialize into object structure format using python numpy pickle class.

**Sentiment in future log return**

By extracting the log return from the next first to fifth day, I plotted it using seaborn regression plotter and obtain the following result.



The plot above shows the result from day 1 log return to the day 3 log return. The slope of the linear regression line decreases from positive slope value in day 1 to nearly zero slope in day 3.



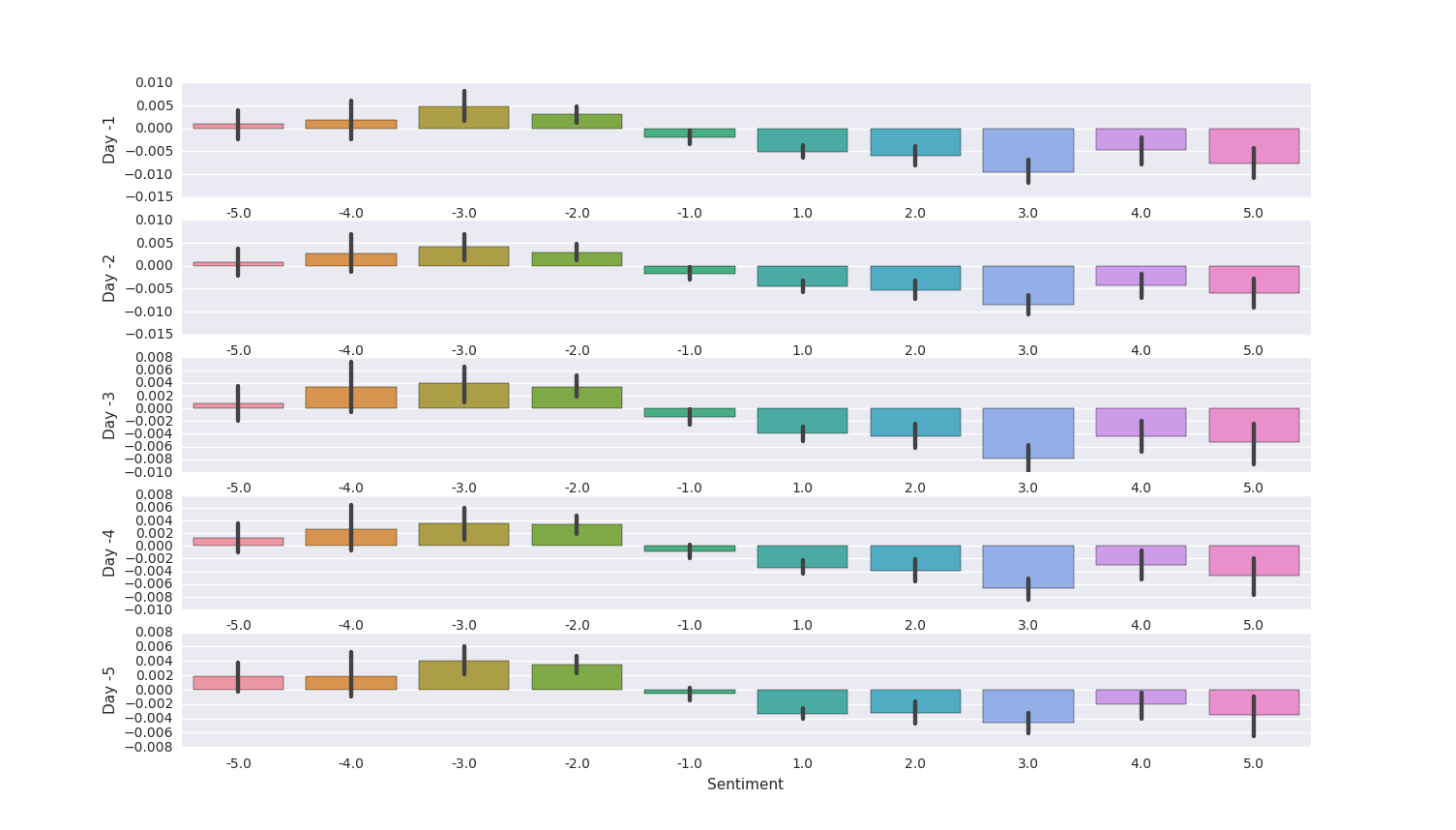
The value of slope continues to decreases in day 4 to day 7. Returning the slope values proven my observation.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Day 6 | Day 7 |
| Slope value | 2.842572 | 0.931753 | 0.272668 | -0.191817 | -0.830956 | -0.852299 | -0.847485 |
| Pearson correlation coefficient | 0.027596 | 0.011580 | 0.003983 | -0.003159 | -0.014979 | -0.016472 | -0.017409 |

Hence, the value of sentiment data only effect on the following day market. The Pearson correlation coefficient further backend my   
assumptions

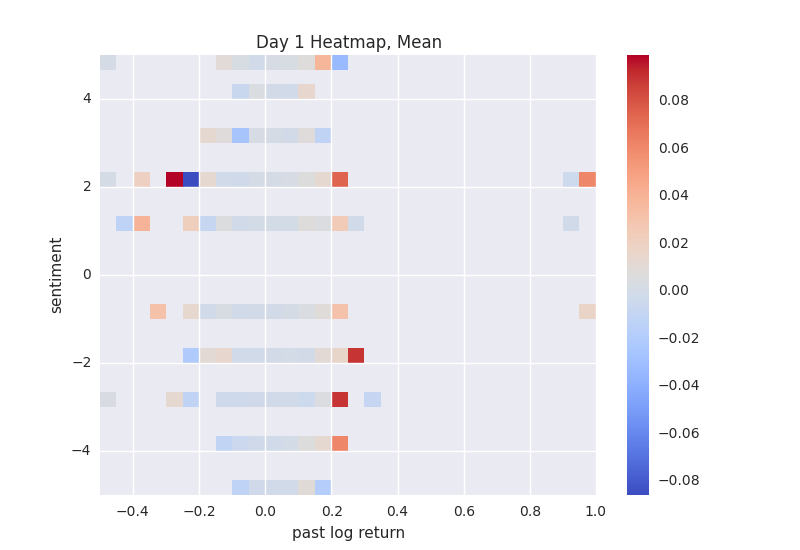
**History log return and sentiment value**

The below plot shows the past log return on the sentiment value. The y axis represent the past 5 days log return value and the x axis showing the sentiment value ranging from -5 to 5. The fluctuation of the log return is denote on the bar line.

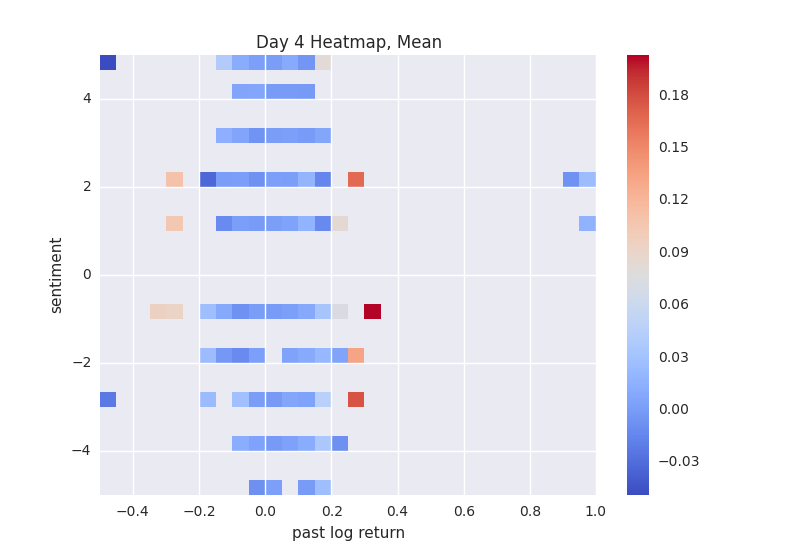
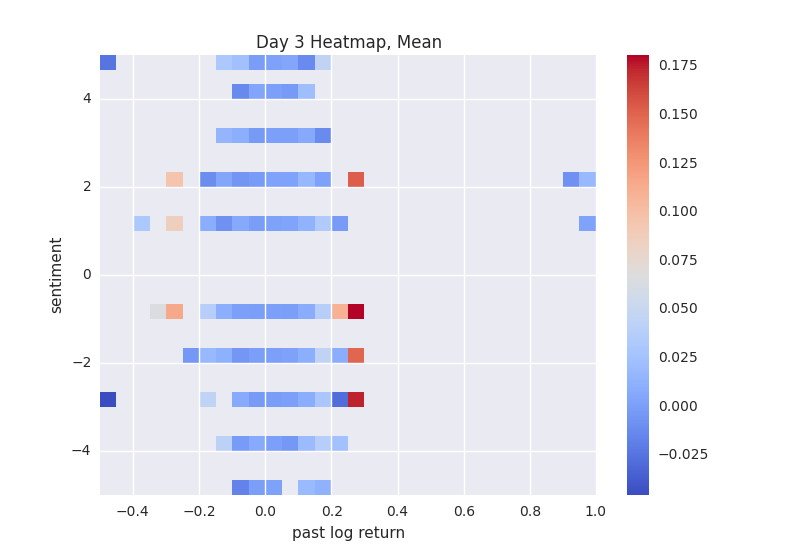
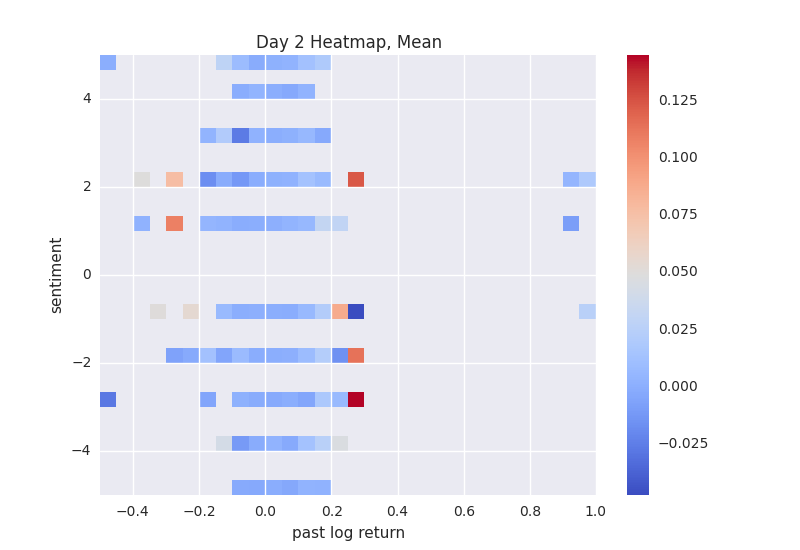


An interesting observation is found in the plot above is the values of past log return has a reverse impact on the sentiment value. Hence, I binned the sentiment values, past log returns and future log returns into a histogram 2d data with future log returns as the counts of bins in each point. Median value is compute for each values in the bins.

The 2d data is visualize into a heatmap plot using python matplotlib imshow function.



An interesting observation was found in the lower right of the points cluster showing that it’s likely to have a positive log return in the following day if there is a positive log return yesterday and a negative sentiment signal today.



The plot result from log return day 2 to day 4(left to right), indicates negative log return as shown in our plot above.

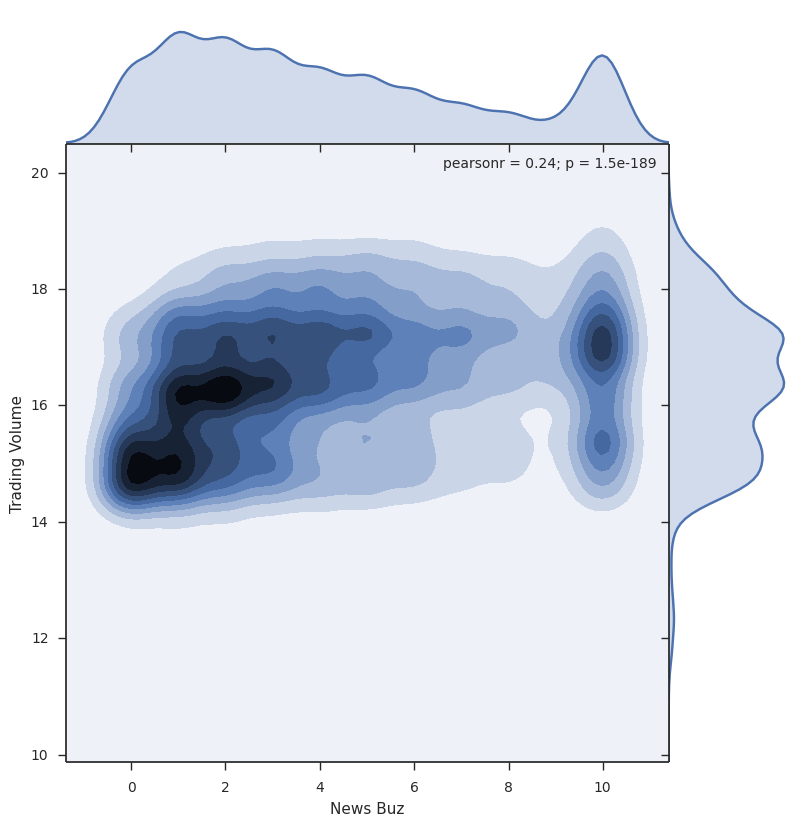
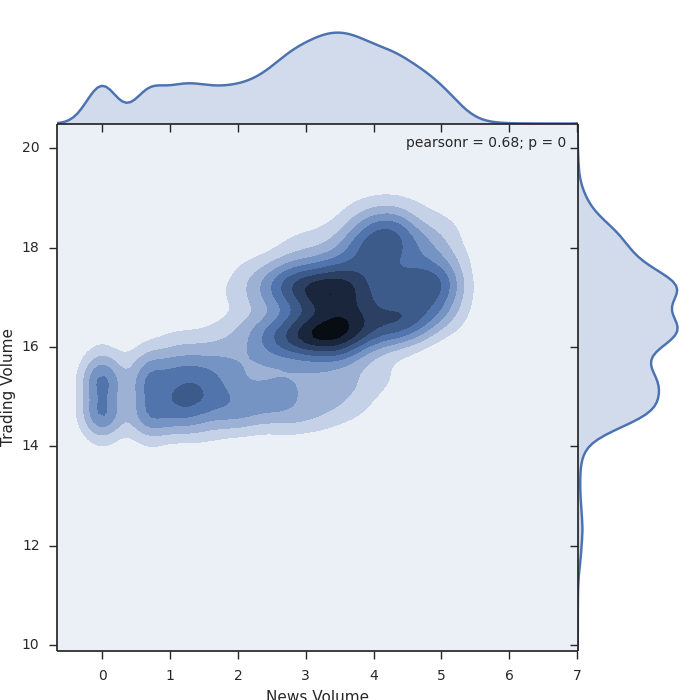
However, the heatmap heat points do not cover the whole plot due to the small database of 14300 data. Hence, if I am given the chance to be an intern in InfoTrie, I wish to further study my observation with a larger database.

**Correlation of Trading volume and News Data**

Looking through Finsents sentiment, I stumble upon the news buzz and news volume columns. I decided to find if the news buzz and news volume affect the trading volumes.

Due to size of daily trading volume are around tens of thousands to millions of trades each day, I also log these data into a more comparable metrics.

The relation between news buzz and news volume are visualize using python’s seaborn jointplot. With denser color indicates the density of the points.



Both plots show a high density of trading volume of 16 react to news volume in news buzz. Indicates if the news volume of a stock reaches critical volume of 3 the trading volume increase from an interval between 14-16 to an interval of 16-18.

News Buzz however, did not show any observable trend as the interval of trading volume are evenly distributed in a range of 14-18.

Source Code

<https://github.com/DBlackKat/Sentiment-Data-Analysis>

The plot result of are found in plot.py files while