**EXECUTIVE SUMMARY**

**Introduction:** Social media has become an essential part of our lives, and with the proliferation of multivariate ideas and opinion circulating these websites, there exists an immense opportunity for us to analyze how consumers impact financial markets with their words. Even the recent trends suggest towards an increase in the impact of digital opinions on consumer decisions. “Deloitte (2007), for instance, finds that 82% of US Internet consumers report to be directly influenced by peer reviews in their purchasing decisions” [[[1]](#footnote-1)]. Having such studies in hand, it may safely be assumed that investors looking to increase the returns from the stock market may be swayed by the opinions and news appearing in relation thereto.

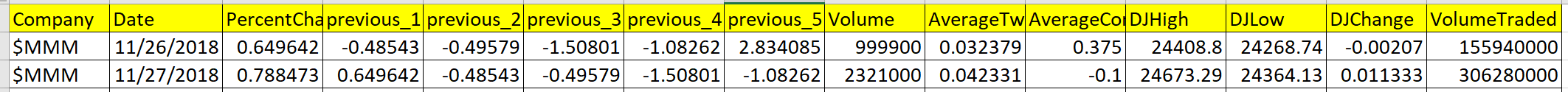
**Procedure:** To examine the impact of social media sentiments on financial markets, we selected the 30 companies falling within the Dow Jones index and extracted the movements in their stock prices over a one-week period. Data (Exhibit 1 and 2) was obtained from the Yahoo Finance website, along with the Dow Jones Industrial Average. We derived the percentage change in the stock prices of individual companies and industrial averages, which constituted the financial aspect of model and reflected on the performance of companies and overall market.

Furthermore, to understand the general market sentiments on the individual companies’ stocks, we extracted the latest tweets and replies available on Twitter. The performance of natural language processing and sentiment analysis on such consumer data was done using hive ‘sentimentanalysis’ followed by comparing scores in LM and General Dictionary. We also compared the scores obtained through LM and AFINN .LM calculated the sentiments more accurately as it has dictionary values for finance related terms which is not present in AFINN dictionary. (Exhibit 3) This allowed us to assign a positive / negative score on the tweets and replies. Such scores reflect the general market sentiment towards a company’s stock and constituted the independent variable in our model.

**Insights and Conclusions:** The visualization in Exhibit 4 shows the percentage change in stock process along with the previous day’s sentiment scores of tweets and corresponding replies. The plots appear to show relationship between the sentiment score and momentum of stocks. Linear regression model was used to confirm the insights.

Percentage change in stock price was chosen as the dependent variable while a combination of various input features such as previous day’s percentage change in stock prices, sentiment scores and Dow Jones averages were also included in the model. Including company and the previous stock changes in model helped account for the variation among companies and the time series factor. Results (Exhibit 5) showed that the sentiment scores of tweets were slightly influencing the momentum of stocks. Since, there is scope for better analysis of sentiment by utilizing the information available in other platforms such as Facebook and reddit, we conclude that tweets or posts on social media plays a significant role in influencing the stock prices.

Investors expect to earn high capital returns while minimizing risks. Investing in a diverse portfolio of stocks helps the investor by minimizing the overall risk since the likelihood of all stocks in a portfolio performing badly in reaction to market events is low. Having said so, it must be noted that current day investors are looking for ways to be savvy in terms of social media. Hence, it is essential to harness the power of social media and apply it within your own business context. Our sentiment analysis aims to do the same by building a portfolio of profitable companies to help recommend more beneficial stocks to our clients. Such analysis will help us make better recommendations for stock market investments and generate revenues by charging per transaction commissions and profit margins.

**APPENDIX**

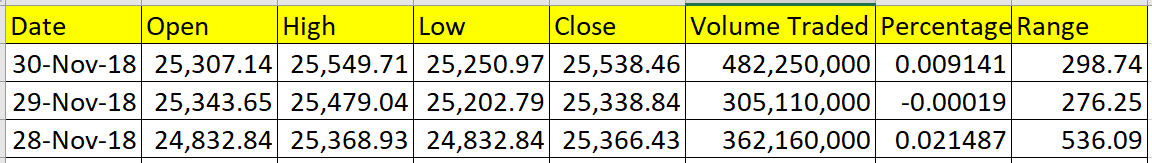


Exhibit 1: Data for Model

Exhibit 2 – DowJones Averages

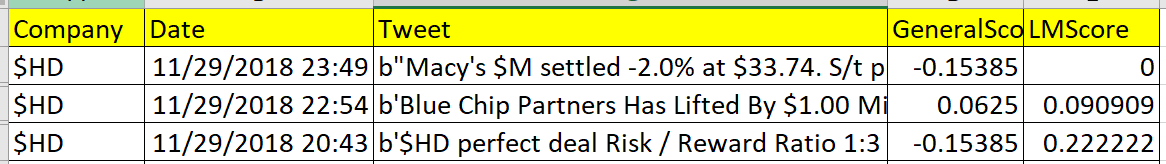


Exhibit 3 – 30 Company Stock Values

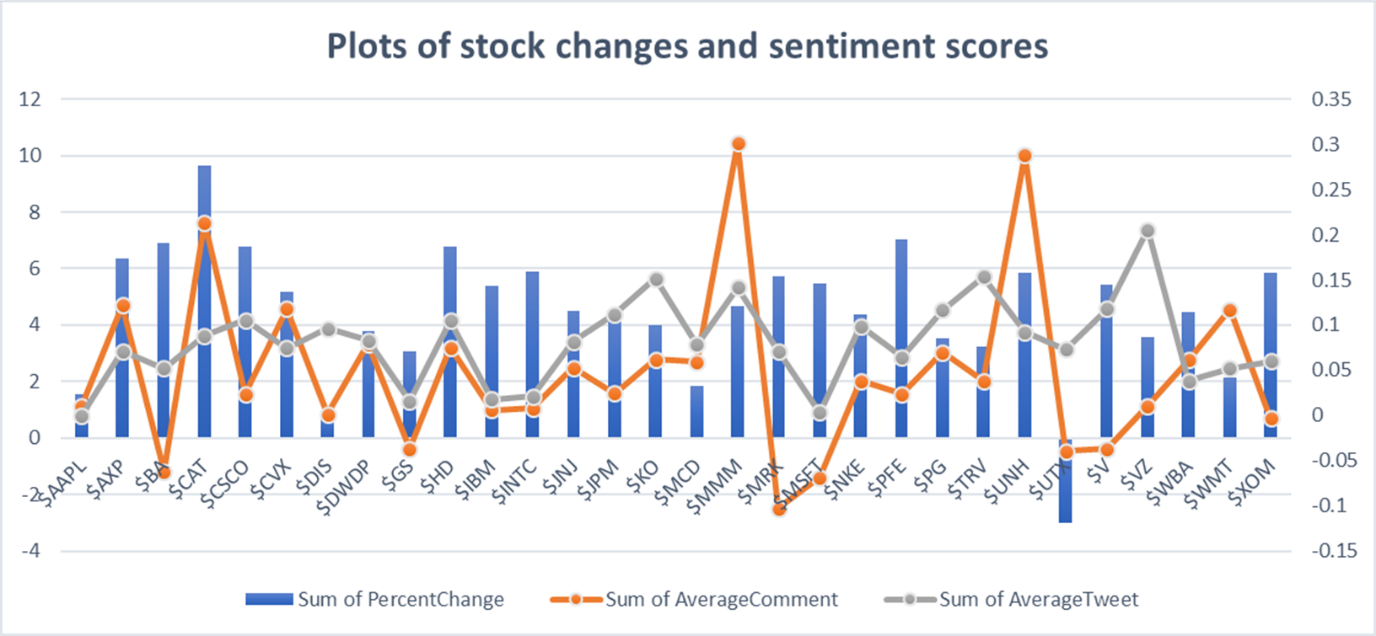
****

Exhibit 4 – Stock Price Changes – Sentiment Score

Exhibit 5 – Linear Regression Model Results

Residuals:

Min 1Q Median 3Q Max

-2.05443 -0.57350 -0.02833 0.59598 1.86035

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.835e+01 1.155e+01 5.050 1.71e-06 \*\*\*

Company$AXP 1.049e+00 6.131e-01 1.711 0.089770 .

Company$BA 1.177e+00 6.081e-01 1.935 0.055485 .

Company$CAT 1.766e+00 6.235e-01 2.833 0.005468 \*\*

Company$CSCO 1.120e+00 6.259e-01 1.789 0.076313 .

Company$CVX 7.406e-01 6.119e-01 1.210 0.228689

Company$DIS -2.993e-01 6.154e-01 -0.486 0.627703

Company$DWDP 3.778e-01 6.124e-01 0.617 0.538493

Company$GS 3.045e-01 5.995e-01 0.508 0.612482

Company$HD 1.011e+00 6.216e-01 1.627 0.106498

Company$IBM 8.315e-01 5.999e-01 1.386 0.168480

Company$INTC 9.429e-01 6.004e-01 1.570 0.119111

Company$JNJ 5.906e-01 6.133e-01 0.963 0.337554

Company$JPM 4.575e-01 6.224e-01 0.735 0.463855

Company$KO 2.055e-01 6.386e-01 0.322 0.748174

Company$MCD -7.116e-02 6.099e-01 -0.117 0.907318

Company$MMM 4.592e-01 6.425e-01 0.715 0.476330

Company$MRK 8.840e-01 6.130e-01 1.442 0.152018

Company$MSFT 1.099e+00 6.063e-01 1.813 0.072468 .

Company$NKE 4.391e-01 6.168e-01 0.712 0.477981

Company$PFE 1.258e+00 6.126e-01 2.054 0.042326 \*

Company$PG 2.393e-01 6.238e-01 0.384 0.702018

Company$TRV 1.061e-01 6.405e-01 0.166 0.868742

Company$UNH 1.096e+00 6.326e-01 1.733 0.085828 .

Company$UTX -1.190e+00 6.083e-01 -1.956 0.052953 .

Company$V 6.794e-01 6.261e-01 1.085 0.280166

Company$VZ 9.243e-03 6.699e-01 0.014 0.989015

Company$WBA 8.841e-01 6.114e-01 1.446 0.150937

Company$WMT 1.452e-01 6.060e-01 0.240 0.811027

Company$XOM 9.043e-01 6.077e-01 1.488 0.139538

previous\_1 -3.237e-01 8.868e-02 -3.650 0.000399 \*\*\*

AverageComment -8.813e-01 1.680e+00 -0.524 0.600994

AverageTweet 1.113e+01 7.254e+00 1.534 0.127750

DJHigh 8.298e-03 3.798e-03 2.185 0.030989 \*

DJLow -1.087e-02 4.186e-03 -2.596 0.010680 \*

DJChange -2.086e+02 4.985e+01 -4.185 5.66e-05 \*\*\*

VolumeTraded 1.473e-08 2.706e-09 5.443 3.08e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9465 on 113 degrees of freedom

Multiple R-squared: 0.4825, Adjusted R-squared: 0.3177

F-statistic: 2.927 on 36 and 113 DF, p-value: 8.28e-06

1. [↑](#footnote-ref-1)