# Social media sentiment analysis of listed companies and its impact on Stock Prices

Francis Kurian

12/12/2021

## Abstract

Social networking sites and news feeds generate voluminous text information that reveal common sentiments about a business. Quantifying sentiments and understanding the relevance of such information on a company's stock price has been an area of interest for many researchers. Sentiment analysis, using extensive text mining algorithms, is technically challenging to build but services like IBM Watson Application Programming Interfaces (APIs) simplify such tasks. This paper examines how stock price movements are correlated with the expressed public sentiments for a select number of companies. To achieve this goal, a tool that integrates relevant data sources and automates the analysis was developed using programming language R. A multiple regression model was employed to quantify the relationship between stock prices and a sentiment index. In addition to the sentiment index, exchange rates were introduced as explanatory variables to capture the international exposure of the companies. Validation of the model on two listed companies in the United States stock exchanges demonstrate significant explanatory power of sentiments index and thus supports the hypothesis that sentiment analysis is a useful predictor of stock price movement.

## Introduction

### Sentiment Analysis and Stock Prices

Consumer behavior studies tell us that emotions can profoundly affect individual behavior and decision-making. Does this also apply to societies at large, i.e. can societies experience mood states that affect their collective decision making when it comes to pricing company stock?. [1] Efficient Market Hypothesis implies that stock market prices incorporate the new information available to the market participants instantaneously. News feeds and social media interactions generate large volume of content in the form of product reviews, service levels, management decisions and many other relevant information that are sensitive to the stock prices. For example, many customers posting negative reviews about a product in social media could impact the product demand and future profitability of a company. News about supply chain bottlenecks for certain electronics parts could signal future miss in in sales targets.

It's anticipated that sentiment analysis is relevant for understanding the stock price movement[2]. Quantification of human sentiments expressed in text documents like news feed or social media interaction is a computationally intensive task.This is where Watson API services help to look at these voluminous documents to capture feelings, such as anger, sadness, joy, fear, hesitation etc. Sentiment is a higher-level classifier that divides the spectrum of emotions into positive, negative, and neutral. Once such classification is done it is possible to build an index to track the sentiments movement on a daily basis and analyse its sensitivity to stock prices. This system leverages on an index already built by TRaiCE Inc using Watson APIs and is available as a service. In following sections will discuss the methodology, data sources and the analysis results.

News about political instability and changes in international business environments could impact the foreign exchange rates and affect the profitability of companies doing business with those nations. When news and social media interactions are in a language other than English, introducing exchange rate as a proxy to the model is one way to indirectly measure the sentiments.

### Quantification of Stock Market Behavior using Sentiment Analysis

Several researchers have empirically established the links between sentiments analysis and stock price movements. Bollen et al[1]. conducted a conclusive study using microblogs from twitter and established a causal relationship with sentiment score and Dow Jones Industrial Average. Seki et al. [3]. Zhang et al.[4] analyzed news text to build a business sentiment index to use it as an economic index in predicting stock prices. Yahya et al.[5] used a linear regression approach incorporating sentiment analysis as an explanatory variable and concluded that introducing sentiment analysis provided significant lift in predicting the stock price movement. On similar lines, a rudimentary regression-based analysis system was developed to understand the stock price movement with respect to business sentiment index (BSI).

### Objectives

This paper attempts to analyze stock price behavior of two companies-The Boeing and Kraft Heinz- listed in the US stock market with respect to changes in Business Sentiment index (BSI) . BSI quantifies the human feelings captured through the extensive text-based content from news feeds and social media interactions. In addition to BSI, foreign exchange rates are added as an additional explanatory variables. This is with the assumption that exchange rates are a better proxy when news and social media interactions are conducted using a language other than English.

## Methodology and Data

### Analysis System and Linear Regression Model

An R based analysis system was developed to integrate stock prices, BSI and various exchange rates at the company level. The design was in such a way that analyst could make a function call with ticker symbol and produce data diagnostics and multiple regression analysis. This study leverages on that system.

Ordinary Least Square Regression:

$X_i$ are the $k$ independent variables and $Y$ is a dependent variable, for each sample of $n$ , the value of $Y_n$ is:

$$Y_n = \sum_{i=0}^{k} \beta_i X_{ni} + \epsilon_i \tag{1}$$

$\epsilon_i$ =Random Error Term

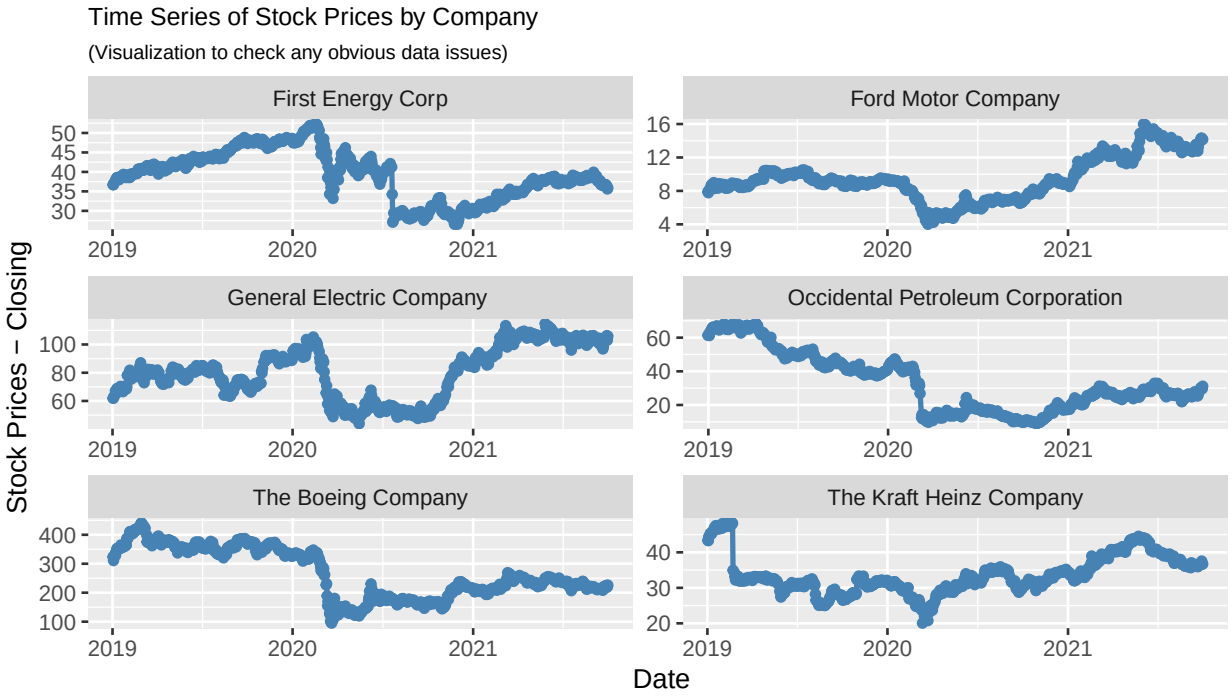
The OLS model in this context can be specified as:

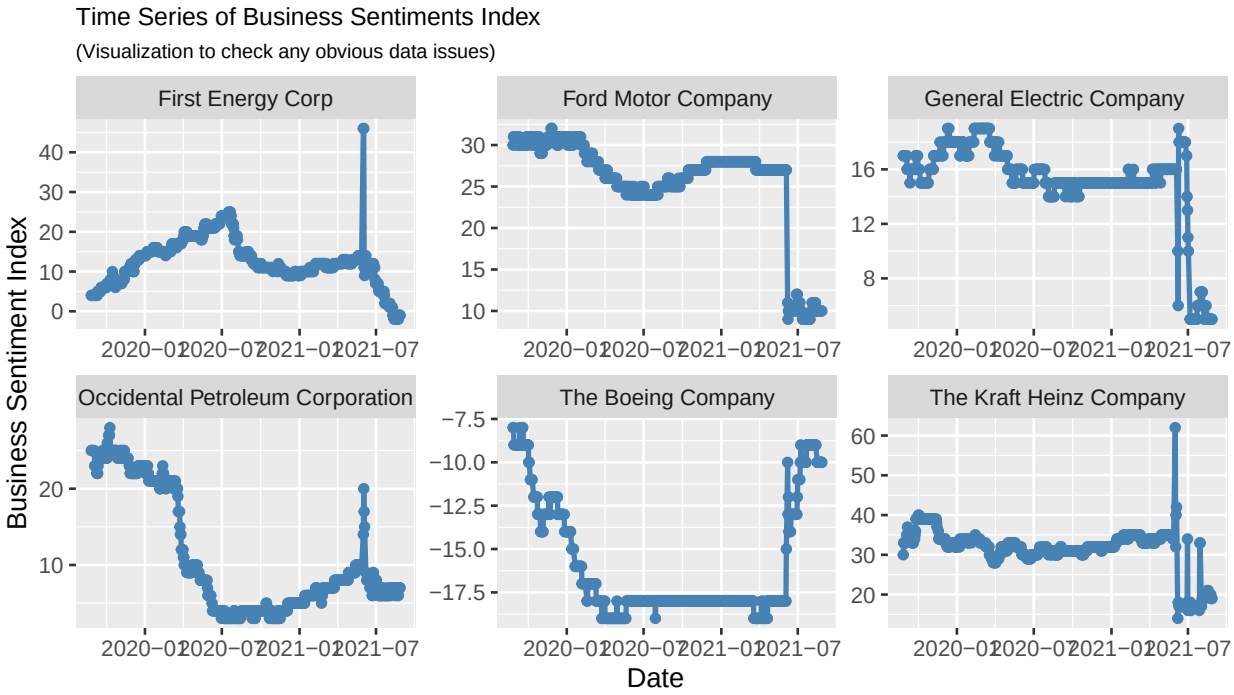
$$StockPrice = \beta_0 + \beta_1 BSI + \beta_2 USD.EUR + \beta_3 USD.YEN + \beta_4 USD.INR \tag{2}$$

where:
**StockPrice** is Stock Price of the company.
**BSI** is Business Sentiment Index.
**USD.EUR** is US Dollar vs Euro exchange rate.
**USD.YEN** is US Dollar vs Japanese Yen exchange rate.
**USD.INR** is US Dollar vs Indian Rupee exchange rate.
$\beta_0$ is the intercept, $\beta_1$,$\beta_2$,$\beta_3$,$\beta_4$ are the coefficients.

### Data Collection and Profiling

Daily data was collected for a period of two years (01-SEP-2019 to 31-AUG-2021) for the following data series. Daily stock prices[6], Business sentiments index[7] and foreign exchange rates[8]. The following graphs show the data distribution for a select number of the companies including the ones analyzed for the study.



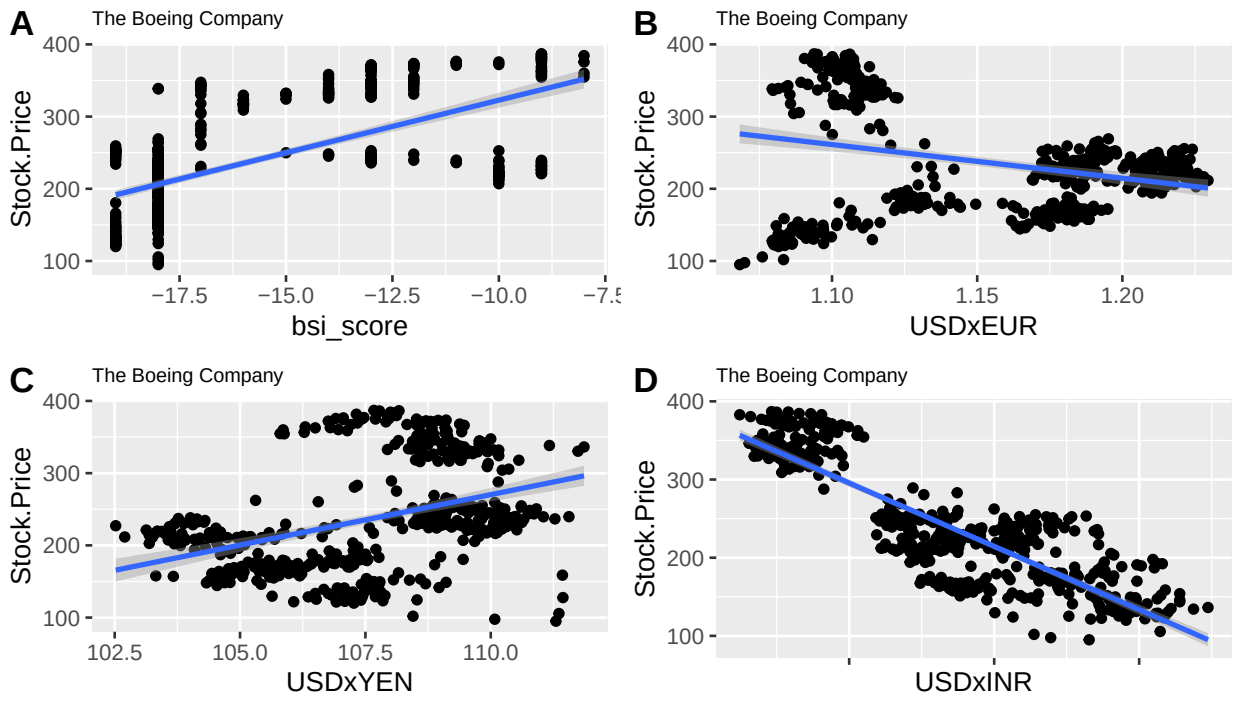

Overall, various time series data demonstrate volatility especially around the early 2020. This could be attributed to the Covid-19 uncertainties in the market at the time. BSI data clearly show some outliers that need a review and probably a correction. With these points noted, the data was used for the further analysis.

### Research Data & Code: Open Science Compliance

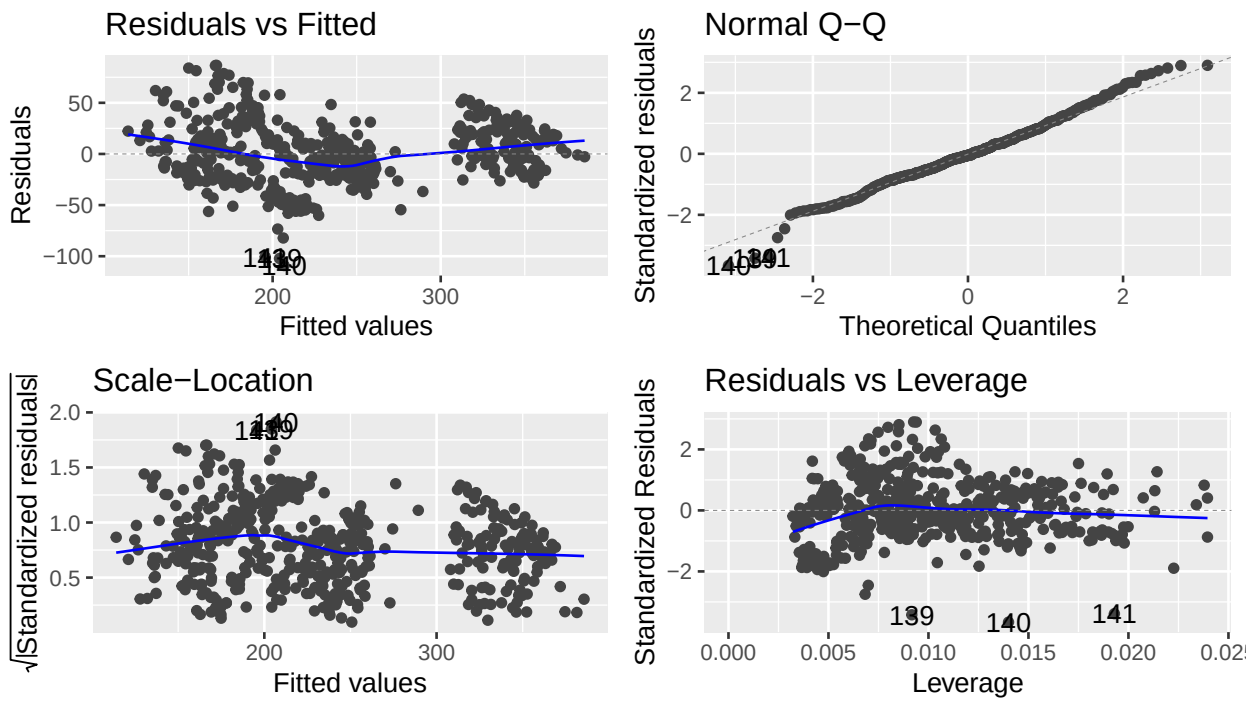
R code developed for data extraction and analysis and data sets used for this validation exercises are archived in a public github repository.Reproducible package can be downloaded using the link provided in the reference section.[9]

## Model Validation and Analysis at Company Level

### The Boeing Company




```
## MULTIPLE REGRESSION RESULTS & ANALYSIS OF ERROR
## Call:
## lm(formula = Stock.Price ~ bsi_score + USDxEUR + USDxYEN + USDxINR,
##     data = df_all)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -109.121  -19.020   -1.475   18.610   86.655
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2411.6876   114.4853  21.065  < 2e-16 ***
## bsi_score      5.4746     0.4967  11.021  < 2e-16 ***
## USDxEUR     -155.1835    30.4394  -5.098 4.90e-07 ***
## USDxYEN        5.1258     0.7355   6.969 1.03e-11 ***
## USDxINR      -33.4920     0.9616 -34.829  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30 on 493 degrees of freedom
## Multiple R-squared:  0.841,  Adjusted R-squared:  0.8398
## F-statistic: 652.1 on 4 and 493 DF,  p-value: < 2.2e-16
```

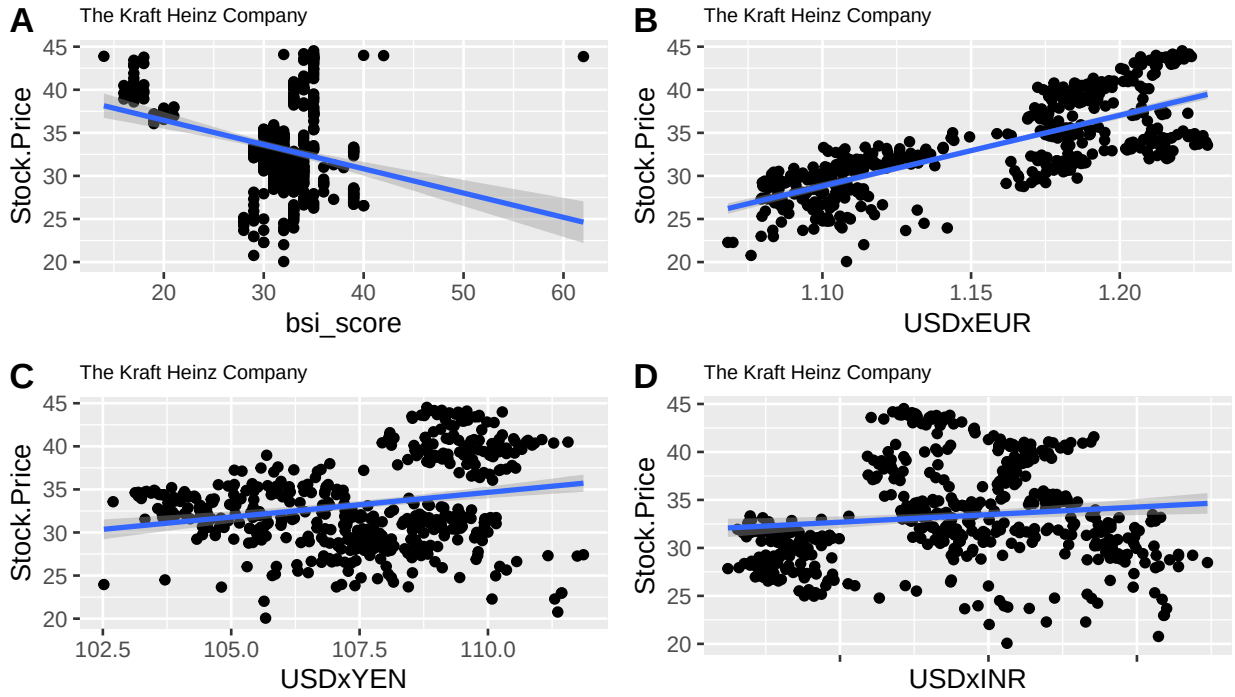



```
## [1] TRUE
```

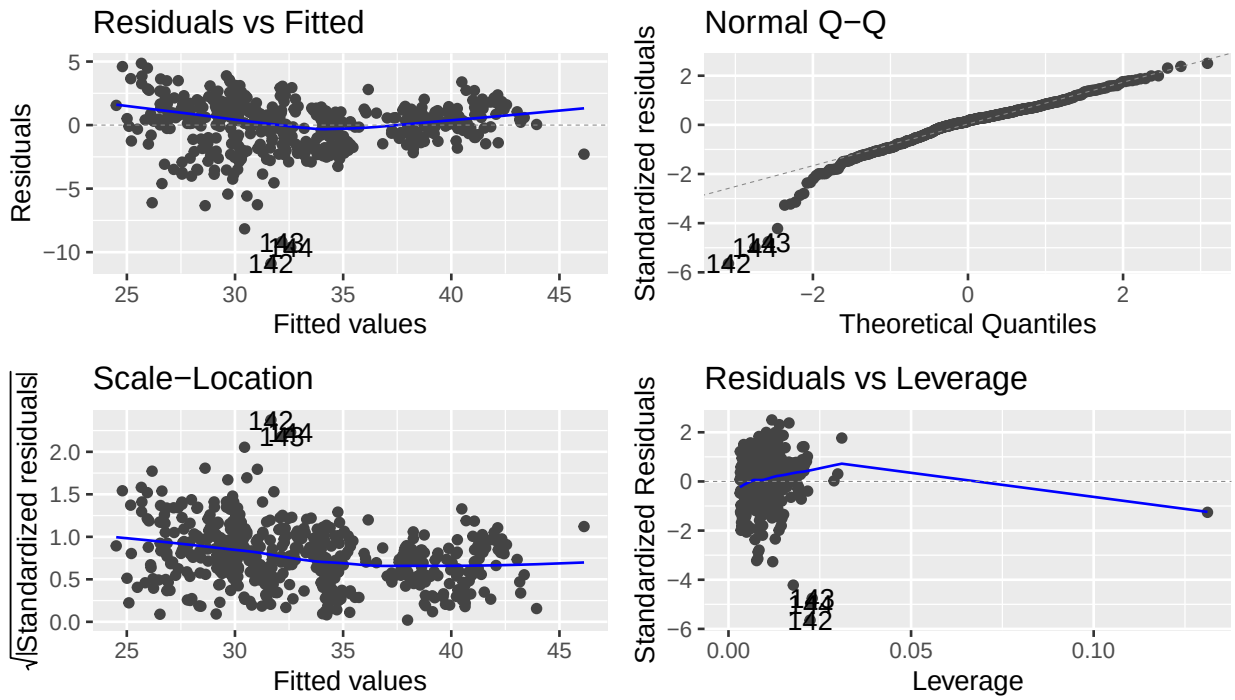
| | Stock Price | | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 2411.69 | 2186.75 – 2636.63 | **<0.001** |
| bsi_score | 5.47 | 4.50 – 6.45 | **<0.001** |
| USDxEUR | -155.18 | -214.99 – -95.38 | **<0.001** |
| USDxYEN | 5.13 | 3.68 – 6.57 | **<0.001** |
| USDxINR | -33.49 | -35.38 – -31.60 | **<0.001** |
| Observations | 498 | | |
| $R^2$ / $R^2$ adjusted | 0.841 / 0.840 | | |

Figure 1: Summary Results of Multiple Regression Analysis

### Kraft Heinz Company




```
## MULTIPLE REGRESSION RESULTS & ANALYSIS OF ERROR
## Call:
## lm(formula = Stock.Price ~ bsi_score + USDxEUR + USDxYEN + USDxINR,
##     data = df_all)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.8938  -1.0429   0.2945   1.1824   4.8579
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -277.68947    9.56027 -29.046  < 2e-16 ***
## bsi_score      0.10314    0.01984   5.200 2.93e-07 ***
## USDxEUR      103.81692    2.09151  49.637  < 2e-16 ***
## USDxYEN        1.44932    0.04949  29.282  < 2e-16 ***
## USDxINR        0.43595    0.06436   6.774 3.58e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.953 on 493 degrees of freedom
## Multiple R-squared:  0.8536, Adjusted R-squared:  0.8525
## F-statistic: 718.8 on 4 and 493 DF,  p-value: < 2.2e-16
```




```
## [1] TRUE
```

| | Stock Price | | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | -277.69 | -296.47 – -258.91 | **<0.001** |
| bsi_score | 0.10 | 0.06 – 0.14 | **<0.001** |
| USDxEUR | 103.82 | 99.71 – 107.93 | **<0.001** |
| USDxYEN | 1.45 | 1.35 – 1.55 | **<0.001** |
| USDxINR | 0.44 | 0.31 – 0.56 | **<0.001** |
| Observations | 498 | | |
| $R^2$ / $R^2$ adjusted | 0.854 / 0.852 | | |

Figure 2: Summary Results of Multiple Regression Analysis

## Summary of Analysis

### The Boeing Company

Boeing is an American multinational corporation that designs, manufactures, and sells airplanes and aviation equipment worldwide. The scatter plots of Boeing stock prices with respect to explanatory variables clearly show a positive relationship with Business Sentiments Index(BSI) and USDxYEN exchange rates. The other two exchanges rates- USDxEUR and USDxINR show a negative relationship.

Multiple regression results and analysis of variance are consistent with scatter plot diagrams. BSI is positively related to stock prices and the coefficient is significant at a 99% confidence interval. This means a unit increase in BSI will lead to 5.5 units increase in the stock prices and vice versa. Euro and Indian Rupee exchange rates are negatively related to stock prices and the $\tau$ statistics is significant at 99% confidence interval levels. One explanation is that high export exposure of Boeing to those regions and a weak foreign currency means higher revenue for the company in terms of dollars.A stronger currency will mean more expensive exports. Japanese Yen on the other hand is positively related to stock prices. Japan's status as a large part supplier to Boeing means imports are favorably sensitive.

Residual analysis provides some more insights into how the random error term in in the equation is distributed. The regression equation $R^2$ square is 84% which means 84% of the stock market price variance is explained by the predictor variables and it's a good fit. So it is evident that sentiments as expressed in news and social media text interactions significantly impact the stock prices.

### Kraft Heinz Company

The Kraft Heinz Company (KHC) is one of the largest food companies in North America. North America is the major market for Heinz products and primary source of raw materials.

Similar to Boeing, coefficient signifcance was measured using $\tau test$. Multiple regression results and analysis of variance show BSI is positively related to stock prices and the coefficient is significant at a 99% confidence interval. Residual analysis shows some inconsistent pattern that require some additional review. All exchange rates are positively related which probably is explained by the company's normalized global exposure. The regression equation $R^2$ is 85% which shows a good fit. So there exists a significant positive relationship between sentiment index and stock prices of Heinz.

## Conclusion and Future Research

Determinants of stock prices is a well-researched area. This paper primarily examined the sensitivity of company's stock price with respect to a business sentiments index. The index quantifies human sentiments expressed in text content generated by news feed or social media interaction about that company. An R based analytical framework was developed for data extraction and analysis and subsequent validation was conducted for two listed companies – The Boeing and Kraft Heinz. The results confirm that there is a significant positive relationship existing between stock prices and human sentiments. The data frequency used was daily so same day sensitivity was measured by the model. This insight can be incorporated into stock trading applications as an additional feature. However, this is a rudimentary approach and scope for improvement include i) refining the model to include lagged variables, additional features and testing the time series properties of the variables. ii) addition of high frequency intra-day data. iii) explore alternate machine learning methods other than the regression.

## References


1. Bollen J, Mao H, Zeng X-J. 2011 Twitter mood predicts the stock market. *Journal of computational science* **2**, 1–8.
2. Mehta P, SharnilPandya, Kotecha K. 2021 Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. *PeerJ Computer Science* **13**.
3. Seki K, Ikuta Y, Matsubayashi Y. 2021 News-based business sentiment and its properties as an economic index. *Information Processing & Management* **59**.
4. Zhang Y, Shirakawa M, Hara T. 2021 Predicting temporary deal success with social media timing signals. *Journal of Intelligent Information Systems*
5. Cakra YE, Trisedya BD. 2015 Stock price prediction using linear regression based on sentiment analysis. *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*
6. *Daily historical data.* https://finance.yahoo.com. Yahoo Finance.
7. Business sentiments index (BSI). *https://www.traice.io*
8. *Foregn exchange rate: Country data.* https://www.federalreserve.gov. Federal Reserve.
9. Kurian F. *Open science compliance: Archived data and r codes to reproduce the results.* https://github.com/FrancisKurian/CS510TermProject edn. GitHub.